A List of Constants

In this appendix, we list all the constants used in our main results, Theorems 3 and 5. They are finite and their expressions do not affect the understanding of the theorems. Since their expressions are quite long and complicated, we begin with the following set of constants, based on which we will be able to present the constants used in the theorems and the proofs of the theorems in an easier way. We hope that this way can also help the readers to better understand and follow our results and analyses.

The first constant ζ_1 is defined as follows. Recall that ϵ is given in (6) as

$$\epsilon = \left(1 + \frac{2b_{\max}}{A_{\max}} - \frac{\pi_{\min}\beta^{2L}}{2\delta_{\max}}\right)(1+\alpha A_{\max})^{2L} - \frac{2b_{\max}}{A_{\max}}(1+\alpha A_{\max})^{L}.$$

ζ_1 is defined as the unique solution for which $\epsilon = 1$ if $\alpha = \zeta_1$. The following remark shows why ζ_1 uniquely exists.

Remark 5 *From* (6)*, it is easy to see that ϵ is monotonically increasing for $\alpha > 0$. Define the corresponding monotonic function as*

$$f(\alpha) = \left(1 + \frac{2b_{\max}}{A_{\max}} - \frac{\pi_{\min}\beta^{2L}}{2\delta_{\max}}\right)(1+\alpha A_{\max})^{2L} - \frac{2b_{\max}}{A_{\max}}(1+\alpha A_{\max})^{L}.$$

Note that $0 < f(0) < 1$ and $f(+\infty) = +\infty$. Thus, $f(\alpha) = 1$ has a unique solution ζ_1. □

The other constants are defined as follows:

$$\zeta_2 = \frac{4b_{\max}^2}{A_{\max}^2}\left[(1+\alpha A_{\max})^L - 1\right]^2 + 2b_{\max}\frac{(1+\alpha A_{\max})^L - 1}{A_{\max}}(1+\alpha A_{\max})^L \tag{12}$$

$$\begin{aligned}\zeta_3 = {} & \left(144 + 4A_{\max}^2 + 912\tau(\alpha)A_{\max}^2 + 168\tau(\alpha)A_{\max}b_{\max}\right)\|\theta^*\|_2^2 \\ & + \tau(\alpha)A_{\max}^2\left[152\left(\frac{b_{\max}}{A_{\max}} + \|\theta^*\|_2\right)^2 + \frac{48b_{\max}}{A_{\max}}\left(\frac{b_{\max}}{A_{\max}} + 1\right)^2 + \frac{87b_{\max}^2}{A_{\max}^2} + \frac{12b_{\max}}{A_{\max}}\right] \\ & + 2 + 2b_{\max}^2 + 4\|\theta^*\|_2^2 + \frac{48b_{\max}^2}{A_{\max}^2}\end{aligned} \tag{13}$$

$$\zeta_4 = \sqrt{N}b_{\max}\left(2 + \frac{12b_{\max}^2}{A_{\max}^2} + 38\|\theta^*\|_2^2\right) \tag{14}$$

$$\zeta_5 = 144 + 916A_{\max}^2 + 168A_{\max}b_{\max} \tag{15}$$

$$\zeta_6 = 4b_{\max}^2\alpha L^2(1+\alpha A_{\max})^{2L-2} + 2b_{\max}L(1+\alpha A_{\max})^{2L-1} \tag{16}$$

$$\begin{aligned}\zeta_7 = {} & (148 + 916A_{\max}^2 + 168A_{\max}b_{\max})\|\theta^*\|_2^2 + 2 + \frac{48b_{\max}^2}{A_{\max}^2} + 152\left(b_{\max} + A_{\max}\|\theta^*\|_2\right)^2 \\ & + 89b_{\max}^2 + 12A_{\max}b_{\max} + 48A_{\max}b_{\max}\left(\frac{b_{\max}}{A_{\max}} + 1\right)^2\end{aligned} \tag{17}$$

$$\zeta_8 = 144 + 916A_{\max}^2 + 168A_{\max}b_{\max} + 144A_{\max}\mu_{\max} \tag{18}$$

$$\begin{aligned}\zeta_9 = {} & \Bigg[2 + (4+\zeta_8)\|\theta^*\|_2^2 + 48\frac{(b_{\max}+\mu_{\max})^2}{A_{\max}^2} + 152\left(b_{\max} + \mu_{\max} + A_{\max}\|\theta^*\|_2\right)^2 \\ & + 12A_{\max}b_{\max} + 48A_{\max}(b_{\max}+\mu_{\max})\left(\frac{b_{\max}+\mu_{\max}}{A_{\max}} + 1\right)^2 + 89(b_{\max}+\mu_{\max})^2\Bigg]\end{aligned} \tag{19}$$

Here $\mu_{\max} = (N+1)A_{\max}C_\theta$, where C_θ is a finite number defined in Lemma 18 which can be regarded as an upper bound of 2-norm of each agent i's state θ_t^i generated by the Push-SA algorithm (9).

A.1 Constants used in Theorem 3

$$
\begin{aligned}
K_1 &= \min\left\{\zeta_1,\ \frac{\gamma_{\max}}{0.9}\right\} \\
K_2 &= 144 + 4A_{\max}^2 + 912\tau(\alpha)A_{\max}^2 + 168\tau(\alpha)A_{\max}b_{\max} \\
C_1 &= \frac{\gamma_{\max}}{\gamma_{\min}}\left(8\exp\left\{2\alpha A_{\max}T_1\right\} + 4\right)\mathbf{E}\left[\|\langle\theta\rangle_0 - \theta^*\|_2^2\right] \\
&\quad + 8\frac{\gamma_{\max}}{\gamma_{\min}}\exp\left\{2\alpha A_{\max}T_1\right\}\left(\|\theta^*\|_2 + \frac{b_{\max}}{A_{\max}}\right)^2 \\
C_2 &= \frac{2\zeta_2}{1-\epsilon} + \frac{\gamma_{\max}}{\gamma_{\min}}\cdot\frac{2\alpha\zeta_3\gamma_{\max} + 4\gamma_{\max}\zeta_4}{0.9} \\
C_3 &= \frac{2\zeta_6}{1-\epsilon} \\
C_4 &= 2\zeta_7\alpha_0 C\frac{\gamma_{\max}}{\gamma_{\min}} \\
C_5 &= 2\alpha_0\zeta_4\frac{\gamma_{\max}}{\gamma_{\min}} \\
C_6 &= 2LT_2\frac{\gamma_{\max}}{\gamma_{\min}}\mathbf{E}\left[\|\langle\theta\rangle_{LT_2} - \theta^*\|_2^2\right]
\end{aligned}
\tag{20}
$$

T_1 is any positive integer such that for all $t \geq T_1$, there hold $t \geq \tau(\alpha)$ and $36\sqrt{N}b_{\max}\eta_{t+1}\gamma_{\max} + K_2\alpha\gamma_{\max} \leq 0.1$.

Remark 6 *We show that T_1 must exist. From $0 < \alpha < \min\{K_1,\ \frac{\log 2}{A_{\max}\tau(\alpha)},\ \frac{0.1}{K_2\gamma_{\max}}\}$, it is easy to see that the feasible set of α is nonempty and $K_2\alpha\gamma_{\max} < 0.1$. Since $\lim_{t\to\infty}\eta_t = 0$ by Lemma 8 and $\tau(\alpha) \leq -C\log\alpha$ by Assumption 3, there exists a time instant $T \geq -C\log\alpha$ such that for any $t \geq T$, there hold $t \geq \tau(\alpha)$ and $\eta_{t+1} \leq (0.1 - K_2\alpha\gamma_{\max})/(36\sqrt{N}b_{\max}\gamma_{\max})$, which implies that T_1 exists.* □

T_2 is any positive integer such that for all $t \geq LT_2$, there hold $\alpha_t \leq \alpha$, $2\tau(\alpha_t) \leq t$, $\tau(\alpha_t)\alpha_{t-\tau(\alpha_t)} \leq \min\{\frac{\log 2}{A_{\max}},\ \frac{0.1}{\zeta_5\gamma_{\max}}\}$ and $\zeta_5\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max} + 36\sqrt{N}b_{\max}\eta_{t+1}\gamma_{\max} \leq 0.1$.

Remark 7 *We explain why T_2 must exist. Since $\alpha_t = \frac{\alpha_0}{t+1}$ is monotonically decreasing for t and $\tau(\alpha_t) \leq -C\log\alpha_t = -C\log\alpha_0 + C\log(t+1)$ from Assumption 3, there exists a positive S_1 such that for any $t \geq S_1$, we have $\alpha_t \leq \alpha$ and $t \geq 2\tau(\alpha_t)$ for any constant $0 < \alpha < \min\{K_1,\ \frac{\log 2}{A_{\max}\tau(\alpha)},\ \frac{0.1}{K_2\gamma_{\max}}\}$. Moreover, it is easy to show that*

$$
\lim_{t\to\infty} t - \tau(\alpha_t) \geq \lim_{t\to\infty} t + C\log\alpha_0 - C\log(t+1) = +\infty,
$$
$$
\lim_{t\to\infty}\tau(\alpha_t)\alpha_{t-\tau(\alpha_t)} \leq \lim_{t\to\infty}\frac{-C\alpha_0\log\alpha_0 + C\alpha_0\log(t+1)}{t-\tau(\alpha_t)+1} = 0.
$$

Then, there exists a positive S_2 such that for any $t \geq S_2$, we have $\tau(\alpha_t)\alpha_{t-\tau(\alpha_t)} \leq \min\{\frac{\log 2}{A_{\max}},\ \frac{0.1}{\zeta_5\gamma_{\max}}\}$. In addition, since $\lim_{t\to\infty}\eta_t = 0$ from Lemma 8, when $\tau(\alpha_t)\alpha_{t-\tau(\alpha_t)} \leq \frac{0.1}{\zeta_5\gamma_{\max}}$, there exists a positive S_3 such that for any $t \geq S_3$, we have $\eta_{t+1} \leq (0.1 - \zeta_5\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max})/(36\sqrt{N}b_{\max}\gamma_{\max})$. Therefore, T_2 must exist as we can simply set $T_2 = \max\{S_1,\ S_2,\ S_3\}$. □

A.2 Constants used in Theorem 5

$$
\begin{aligned}
C_7 &= \frac{16}{\epsilon_1}\mathbf{E}\left[\Big\|\sum_{i=1}^{N}\tilde{\theta}_0^i + \alpha_0 A(X_0)\tilde{\theta}_0^i + \alpha_0 b^i(X_0)\Big\|_2\right] \\
C_8 &= \frac{16}{\epsilon_1}\cdot\frac{A_{\max}C_\theta + b_{\max}}{1-\bar{\epsilon}} \\
C_9 &= 2A_{\max}C_\theta + 2b_{\max} \\
C_{10} &= 2N\zeta_9\alpha_0 C\frac{\gamma_{\max}}{\gamma_{\min}} \\
C_{11} &= 2\alpha_0 N\frac{\gamma_{\max}}{\gamma_{\min}} \\
C_{12} &= 2\bar{T}N\frac{\gamma_{\max}}{\gamma_{\min}}\mathbf{E}\left[\|\langle\tilde{\theta}\rangle_{\bar{T}} - \theta^*\|_2^2\right]
\end{aligned}
$$

Here ϵ_1 is a positive constant defined as $\epsilon_1 = \inf_{t\geq 0}\min_{i\in\mathcal{V}}(\hat{W}_t\cdots\hat{W}_0\mathbf{1}_N)^i$. From Corollary 2 (b) in [1] and the fact that each $\hat{W}_t$ is column stochastic, $\epsilon_1 \in [\frac{1}{N^{NL}}, 1]$. See Lemma 19 for more details.

$\bar{T}$ is any positive integer such that for all $t \geq \bar{T}$, there hold $2\tau(\alpha_t) \leq t$, $\mu_t + \tau(\alpha_t)\alpha_{t-\tau(\alpha_t)}\zeta_8 \leq \frac{0.1}{\gamma_{\max}}$ and $\tau(\alpha_t)\alpha_{t-\tau(\alpha_t)} \leq \min\{\frac{\log 2}{A_{\max}}, \frac{0.1}{\zeta_8\gamma_{\max}}\}$.

Remark 8 *From Lemma 20,* $\lim_{t\to\infty}\mu_t = 0$. *Then, using the similar arguments as in Remark 7, we can show the existence of* $\bar{T}$. □

B Analysis and Proofs

In this appendix, we provide the analysis of our two algorithms, (1) and (9), and the proofs of all the assertions in the paper. We begin with some notation.

B.1 Notation

We use $\mathbf{0}_n$ to denote the vector in $\mathbb{R}^n$ whose entries all equal to 0's. For any vector $x \in \mathbb{R}^n$, we use $\operatorname{diag}(x)$ to denote the $n\times n$ diagonal matrix whose ith diagonal entry equals x^i. We use $\|\cdot\|_F$ to denote the Frobenius norm. For any positive diagonal matrix $W \in \mathbb{R}^{n\times n}$, we use $\|A\|_W$ to denote the weighted Frobenius norm for $A \in \mathbb{R}^{n\times m}$, defined as $\|A\|_W = \|W^{\frac{1}{2}}A\|_F$. It is easy to see that $\|\cdot\|_W$ is a matrix norm. We use $\mathbf{P}(\cdot)$ to denote the probability of an event and $\mathbf{E}(X)$ to denote the expected value of a random variable X.

B.2 Distributed Stochastic Approximation

In this subsection, we analyze the distributed stochastic approximation algorithm (1) and provide the proofs of the results in Section 2. We begin with the asymptotic performance.

Proof of Lemma 1: Since the uniformly strongly connectedness is equivalent to B-connectedness as discussed in Remark 2, the existence is proved in Lemma 5.8 of [2], and the uniqueness is proved in Lemma 1 of [3]. ■

Proof of Theorem 1: Without loss of generality, let $\{\mathbb{G}_t\}$ be uniformly strongly connected by sub-sequences of length L. Note that for any $i \in \mathcal{V}$, we have

$$
0 \leq \pi_{\min}\|\theta_t^i - \langle\theta\rangle_t\|_2^2 \leq \pi_{\min}\sum_{j=1}^{N}\|\theta_t^j - \langle\theta\rangle_t\|_2^2 \leq \sum_{j=1}^{N}\pi_t^j\|\theta_t^j - \langle\theta\rangle_t\|_2^2, \tag{21}
$$

where $\pi_{\min}$ is defined in Lemma 1. From Lemma 9,

$$\begin{aligned}
&\lim_{t\to\infty}\sum_{i=1}^{N}\pi_t^i\|\theta_t^i-\langle\theta\rangle_t\|_2^2\\
&\le \lim_{t\to\infty}\hat{\epsilon}^{q_t-T_4^*}\sum_{i=1}^{N}\pi^i_{T_4^*L+m_t}\|\theta^i_{T_4^*L+m_t}-\langle\theta\rangle_{T_4^*L+m_t}\|_2^2+\lim_{t\to\infty}\frac{\zeta_6}{1-\hat{\epsilon}}\left(\alpha_0\hat{\epsilon}^{\frac{q_t-1}{2}}+\alpha_{\lceil\frac{q_t-1}{2}\rceil L}\right)\\
&=0. \qquad (22)
\end{aligned}$$

Combining (21) and (22), it follows that for all $i\in\mathcal{V}$, $\lim_{t\to\infty}\pi_{\min}\|\theta_t^i-\langle\theta\rangle_t\|_2^2=0$. Since $\pi_{\min}>0$ by Lemma 1, $\lim_{t\to\infty}\|\theta_t^i-\langle\theta\rangle_t\|_2=0$ for all $i\in\mathcal{V}$. ■

Proof of Theorem 2: From Theorem 1, all θ_t^i, $i\in\mathcal{V}$, will reach a consensus with $\langle\theta\rangle_t$ and the update of $\langle\theta\rangle_t$ is given in (4), which can be treated as a single-agent linear stochastic approximation whose corresponding ODE is (5). From [4, 5],[1] we know that $\langle\theta\rangle_t$ will converge to θ^* w.p.1, which implies that θ_t^i will converge to θ^* w.p.1. In addition, from Theorem 3-(2) and Lemma 8, $\lim_{\to\infty}\sum_{i=1}^N\pi_t^i\mathbf{E}[\|\theta_t^i-\theta^*\|_2^2]=0$. Since π_t^i is uniformly bounded below by $\pi_{\min}>0$, as shown in Lemma 1, it follows that θ_t^i will converge to θ^* in mean square for all $i\in\mathcal{V}$. ■

We now analyze the finite-time performance of (1). In the sequel, we use K to denote the dimension of each θ_t^i, i.e., $\theta_t^i\in\mathbb{R}^K$ for all $i\in\mathcal{V}$.

B.2.1 Fixed Step-size

We first consider the fixed step-size case and begin with validation of two "convergence rates" in Theorem 3.

Lemma 2 *Both ϵ and $(1-\frac{0.9\alpha}{\gamma_{\max}})$ lie in the interval $(0,1)$.*

Proof of Lemma 2: Since $0<\alpha<K_1=\min\{\zeta_1,\ \frac{\gamma_{\max}}{0.9}\}$ as imposed in Theorem 3, we have $0<\alpha<\zeta_1$ and $0<\alpha<\frac{\gamma_{\max}}{0.9}$. The latter immediately implies that $1-\frac{0.9\alpha}{\gamma_{\max}}\in(0,1)$. From Remark 5, ϵ is monotonically increasing for $\alpha>0$. In addition, from the definition of ζ_1 in Section A that if $\alpha=\zeta_1$, then $\epsilon=1$. Since $0<\alpha<\zeta_1$, it follows that $0<\epsilon<1$. ■

To proceed, we need the following derivation and lemmas.

Let $Y_t=\Theta_t-\mathbf{1}_N\langle\theta\rangle_t^\top=(I-\mathbf{1}_N\pi_t^\top)\Theta_t$. For any $t\ge s\ge 0$, let $W_{s:t}=W_tW_{t-1}\cdots W_s$. Then,

$$\begin{aligned}
Y_{t+1}&=\Theta_{t+1}-\mathbf{1}_N\langle\theta\rangle_{t+1}^\top\\
&=W_t\Theta_t+\alpha W_t\Theta_tA^\top(X_t)+\alpha B(X_t)-\mathbf{1}_N(\langle\theta\rangle_t^\top+\alpha\langle\theta\rangle_t^\top A^\top(X_t)+\alpha\pi_{t+1}^\top B(X_t))\\
&=W_t(I-\mathbf{1}_N\pi_t^\top)\Theta_t+\alpha W_t(I-\mathbf{1}_N\pi_t^\top)\Theta_tA^\top(X_t)+\alpha(I-\mathbf{1}_N\pi_{t+1}^\top)B(X_t)\\
&=W_tY_t+\alpha W_tY_tA^\top(X_t)+\alpha(I-\mathbf{1}_N\pi_{t+1}^\top)B(X_t). \qquad (23)
\end{aligned}$$

For simplicity, let Y_t^i be the i-th column of matrix $Y_t^\top$. Then,

$$Y_{t+1}^i=\sum_{j=1}^N w_t^{ij}Y_t^j+\alpha A(X_t)\sum_{j=1}^N w_t^{ij}Y_t^j+\alpha\left(b^i(X_t)-B^\top(X_t)\pi_{t+1}\right). \qquad (24)$$

[1] On page 1289 of [4], it says that the idea in [5] can be adapted to get the w.p.1 convergence result.

From (23), we have

$$
\begin{aligned}
Y_{t+L} &= W_{t+L-1}Y_{t+L-1}(I+\alpha A^\top(X_{t+L-1})) + \alpha(I-\mathbf{1}_N\pi_{t+L}^\top)B(X_{t+L-1}) \\
&= W_{t+L-1}W_{t+L-2}Y_{t+L-2}(I+\alpha A^\top(X_{t+L-2}))(I+\alpha A^\top(X_{t+L-1})) \\
&\quad + \alpha W_{t+L-1}(I-\mathbf{1}_N\pi_{t+L-1}^\top)B(X_{t+L-2})(I+\alpha A^\top(X_{t+L-1})) \\
&\quad + \alpha(I-\mathbf{1}_N\pi_{t+L}^\top)B(X_{t+L-1}) \\
&= W_{t:t+L-1}Y_t(I+\alpha A^\top(X_t))\cdots(I+\alpha A^\top(X_{t+L-1})) + \alpha(I-\mathbf{1}_N\pi_{t+L}^\top)B(X_{t+L-1}) \\
&\quad + \alpha\sum_{k=t}^{t+L-2} W_{k+1:t+L-1}(I-\mathbf{1}_N\pi_{k+1}^\top)B(X_k)\left(\Pi_{j=k+1}^{t+L-1}(I+\alpha A^\top(X_j))\right),
\end{aligned} \tag{25}
$$

and

$$
Y_{t+L}^i = \left(\Pi_{k=t}^{t+L-1}(I+\alpha A(X_k))\right)\sum_{j=1}^N w_{t:t+L-1}^{ij}Y_t^j + \alpha\hat{b}_{t+L}^i,
$$

where

$$
\begin{aligned}
\hat{b}_{t+L}^i &= (b^i(X_{t+L-1}) - B(X_{t+L-1})^\top\pi_{t+L}) \\
&\quad + \sum_{k=t}^{t+L-2}\left(\Pi_{j=k+1}^{t+L-1}(I+\alpha A(X_j))\right)\sum_{j=1}^N w_{k+1:t+L-1}^{ij}(b^j(X_k) - B(X_k)^\top\pi_{k+1}).
\end{aligned}
$$

Lemma 3 *Suppose that Assumption 1 holds and $\{\mathbb{G}_t\}$ is uniformly strongly connected by sub-sequences of length L. Then, for all $t \geq 0$,*

$$
\sum_{i=1}^N \pi_{t+L}^i \sum_{j=1}^N\sum_{k=1}^N w_{t:t+L-1}^{ij}w_{t:t+L-1}^{ik}\|Y_t^j - Y_t^k\|_2^2 \geq \frac{\pi_{\min}\beta^{2L}}{\delta_{\max}}\sum_{i=1}^N \pi_t^i\|Y_t^i\|_2^2,
$$

where $\beta > 0$ and $\pi_{\min} > 0$ are given in Assumption 1 and Lemma 1, respectively.

Proof of Lemma 3: We first consider the case when $K = 1$, i.e., $Y_t^i \in \mathbb{R}, \forall i$. From Lemma 1, we have

$$
\begin{aligned}
&\sum_{i=1}^N \pi_{t+L}^i \sum_{j=1}^N\sum_{l=1}^N w_{t:t+L-1}^{ij}w_{t:t+L-1}^{il}\|Y_t^j - Y_t^l\|_2^2 \\
&\geq \pi_{\min}\sum_{i=1}^N\sum_{j=1}^N\sum_{l=1}^N w_{t:t+L-1}^{ij}w_{t:t+L-1}^{il}\|Y_t^j - Y_t^l\|_2^2.
\end{aligned}
$$

Let j^* and l^* be the indices such that

$$
|Y_t^{j^*} - Y_t^{l^*}| = \max_{1\leq j,l\leq N}|Y_t^j - Y_t^l|.
$$

From the definition of Y_t, $Y_t^j - Y_t^l = \theta_t^j - \theta_t^l$ for all $j, l \in \mathcal{V}$, which implies that

$$
|Y_t^{j^*} - Y_t^{l^*}| = \max_{1\leq j,l\leq N}|Y_t^j - Y_t^l| = \max_{1\leq j,l\leq N}|\theta_t^j - \theta_t^l| = |\theta_t^{j^*} - \theta_t^{l^*}|.
$$

Since $\cup_{k=t}^{t+L-1}\mathbb{G}_k$ is a strongly connected graph for all $t \geq 0$, we can find a shortest path from agent j^* to agent l^*: $(j_0, j_1), \cdots, (j_{p-1}, j_p)$ with $j_0 = j^*$, $j_p = l^*$, and (j_{m-1}, j_m) is the edge of graph $\cup_{k=t}^{t+L-1}\mathbb{G}_k$, for $1 \leq m \leq p$, which implies that

$$
\begin{aligned}
&\sum_{i=1}^N\sum_{j=1}^N\sum_{l=1}^N w_{t:t+L-1}^{ij}w_{t:t+L-1}^{il}\|Y_t^j - Y_t^l\|_2^2 \\
&\geq \sum_{i=1}^N\sum_{m=1}^p w_{t:t+L-1}^{ij_{m-1}}w_{t:t+L-1}^{ij_m}(Y_t^{j_{m-1}} - Y_t^{j_m})^2.
\end{aligned} \tag{26}
$$

Moreover, we have

$$\sum_{i=1}^{N} w_{t:t+L-1}^{ij_{m-1}} w_{t:t+L-1}^{ij_m} \geq w_{t:t+L-1}^{j_{m-1}j_{m-1}} w_{t:t+L-1}^{j_{m-1}j_m} + w_{t:t+L-1}^{j_m j_{m-1}} w_{t:t+L-1}^{j_m j_m} \geq \beta^{2L}. \tag{27}$$

Then, from Jensen's inequality, (26) and (27), we have

$$\begin{aligned}
&\sum_{i=1}^{N} \pi_{t+L}^{i} \sum_{j=1}^{N} \sum_{l=1}^{N} w_{t:t+L-1}^{ij} w_{t:t+L-1}^{il} \|Y_t^j - Y_t^l\|_2^2 \\
&\geq \pi_{\min} \sum_{i=1}^{N} \sum_{m=1}^{p} w_{t:t+L-1}^{ij_{m-1}} w_{t:t+L-1}^{ij_m} (Y_t^{j_{m-1}} - Y_t^{j_m})^2 \\
&\geq \frac{\pi_{\min}\beta^{2L}}{p} (Y_t^{j^*} - Y_t^{l^*})^2 = \frac{\pi_{\min}\beta^{2L}}{\delta_t} (\theta_t^{j^*} - \theta_t^{l^*})^2.
\end{aligned} \tag{28}$$

For the case when $K > 1$, let Y_t^{ik} be the k-th entry of vector Y_t^i. Then,

$$\begin{aligned}
&\sum_{i=1}^{N} \pi_{t+L}^{i} \sum_{j=1}^{N} \sum_{l=1}^{N} w_{t:t+L-1}^{ij} w_{t:t+L-1}^{il} \|Y_t^j - Y_t^l\|_2^2 \\
&= \sum_{k=1}^{K} \sum_{i=1}^{N} \pi_{t+L}^{i} \sum_{j=1}^{N} \sum_{l=1}^{N} w_{t:t+L-1}^{ij} w_{t:t+L-1}^{il} (Y_t^{jk} - Y_t^{lk})^2.
\end{aligned}$$

For each entry k, we have

$$\sum_{i=1}^{N} \pi_{t+L}^{i} \sum_{j=1}^{N} \sum_{l=1}^{N} w_{t:t+L-1}^{ij} w_{t:t+L-1}^{il} (Y_t^{jk} - Y_t^{lk})^2 \geq \frac{\pi_{\min}\beta^{2L}}{\delta_{\max}} \max_{1 \leq j,l \leq N} (\theta_t^{jk} - \theta_t^{lk})^2, \tag{29}$$

where θ_t^{ik} is the k-th entry of vector θ_t^i. Moreover, let $\Theta_t^{\cdot k}$ be the k-th column of matrix Θ_t.

Since $2x_1x_2 \leq x_1^2 + x_2^2$, we have for any entry $k = 1, \ldots, K$,

$$\begin{aligned}
\sum_{i=1}^{N} \pi_t^i (Y_t^{ik})^2 &= \sum_{i=1}^{N} \pi_t^i \|\theta_t^{ik} - \pi_t^\top \Theta_t^{\cdot k}\|_2^2 \\
&\leq \max_{1 \leq i \leq N} \left[\theta_t^{ik} - \pi_t^\top \Theta_t^{\cdot k}\right]^2 = \max_{1 \leq i \leq N} \left[\pi_t^\top (\mathbf{1}_N \theta_t^{ik} - \Theta_t^{\cdot k})\right]^2 \\
&= \max_{1 \leq i \leq N} \left[\sum_{j=1}^{N} \pi_t^j (\theta_t^{ik} - \theta_t^{jk})\right]^2 = \max_{1 \leq i \leq N} \sum_{j=1}^{N} \sum_{l=1}^{N} \pi_t^j \pi_t^l (\theta_t^{ik} - \theta_t^{jk})(\theta_t^{ik} - \theta_t^{lk}) \\
&\leq \max_{1 \leq i \leq N} \sum_{j=1}^{N} (\pi_t^j)^2 (\theta_t^{ik} - \theta_t^{jk})^2 \leq \max_{1 \leq i \leq N} \sum_{j=1}^{N} \pi_t^j (\theta_t^{ik} - \theta_t^{jk})^2 \\
&\leq \max_{1 \leq i \leq N} \max_{1 \leq j \leq N} (\theta_t^{ik} - \theta_t^{jk})^2.
\end{aligned}$$

Then, combining this inequality with (28) and (29), we have

$$\begin{aligned}
&\sum_{k=1}^{K} \sum_{i=1}^{N} \pi_{t+L}^{i} \sum_{j=1}^{N} \sum_{l=1}^{N} w_{t:t+L-1}^{ij} w_{t:t+L-1}^{il} (Y_t^{jk} - Y_t^{lk})^2 \\
&\geq \frac{\pi_{\min}\beta^{2L}}{\delta_{\max}} \sum_{k=1}^{K} \max_{1 \leq j,l \leq N} (\theta_t^{jk} - \theta_t^{lk})^2 \\
&= \frac{\pi_{\min}\beta^{2L}}{\delta_{\max}} \sum_{k=1}^{K} \sum_{i=1}^{N} \pi_t^i (Y_t^{ik})^2 = \frac{\pi_{\min}\beta^{2L}}{\delta_{\max}} \sum_{i=1}^{N} \pi_t^i \|Y_t^i\|_2^2.
\end{aligned}$$

This completes the proof. ■

Lemma 4 *Suppose that Assumptions 1 and 2 hold and $\{\mathbb{G}_t\}$ is uniformly strongly connected by sub-sequences of length L. Then, when $\alpha \in (0, \zeta_1)$, we have for all $t \geq \tau(\alpha)$,*

$$\sum_{i=1}^N \pi_t^i \|\theta_t^i - \langle\theta\rangle_t\|_2^2 \leq \epsilon^{q_t} \sum_{i=1}^N \pi_{m_t}^i \|\theta_{m_t}^i - \langle\theta\rangle_{m_t}\|_2^2 + \frac{\zeta_2}{1-\epsilon},$$

where ζ_1 is defined in Appendix A, ϵ and ζ_2 are defined in (6) *and* (12)*, respectively.*

Proof of Lemma 4: Let $M_t = \mathrm{diag}(\pi_t)$. Recall the update of Y_{t+L}^i,

$$Y_{t+L}^i = \left(\Pi_{k=t}^{t+L-1}(I + \alpha A(X_k))\right) \sum_{j=1}^N w_{t:t+L-1}^{ij} Y_t^j + \alpha \hat{b}_{t+L}^i.$$

Then, we have

$$\begin{aligned}
\|Y_{t+L}\|_{M_{t+L}}^2 &= \sum_{i=1}^N \pi_{t+L}^i \|Y_{t+L}^i\|_2^2 \\
&= \sum_{i=1}^N \pi_{t+L}^i \| \left(\Pi_{k=t}^{t+L-1}(I + \alpha A(X_k))\right) \sum_{j=1}^N w_{t:t+L-1}^{ij} Y_t^j \|_2^2 && (30)\\
&\quad + \alpha^2 \sum_{i=1}^N \pi_{t+L}^i \|\hat{b}_{t+L}^i\|_2^2 && (31)\\
&\quad + 2\alpha \sum_{i=1}^N \pi_{t+L}^i (\hat{b}_{t+L}^i)^\top \left(\Pi_{k=t}^{t+L-1}(I + \alpha A(X_k))\right) \sum_{j=1}^N w_{t:t+L-1}^{ij} Y_t^j. && (32)
\end{aligned}$$

For (30), since $2(x_1)^\top x_2 = \|x_1\|_2^2 + \|x_2\|_2^2 - \|x_1 - x_2\|_2^2$ and $\pi_t^\top = \pi_{t+L}^\top W_{t:t+L-1}$, we have

$$\begin{aligned}
&\sum_{i=1}^N \pi_{t+L}^i \| \left(\Pi_{k=t}^{t+L-1}(I + \alpha A(X_k))\right) \sum_{j=1}^N w_{t:t+L-1}^{ij} Y_t^j \|_2^2 \\
&\leq (1 + \alpha A_{\max})^{2L} \sum_{i=1}^N \pi_{t+L}^i \| \sum_{j=1}^N w_{t:t+L-1}^{ij} Y_t^j \|_2^2 \\
&= (1 + \alpha A_{\max})^{2L} \sum_{i=1}^N \pi_{t+L}^i \sum_{j=1}^N \sum_{l=1}^N w_{t:t+L-1}^{ij} w_{t:t+L-1}^{il} \frac{1}{2} \left[\|Y_t^j\|_2^2 + \|Y_t^l\|_2^2 - \|Y_t^j - Y_t^l\|_2^2 \right] \\
&= (1 + \alpha A_{\max})^{2L} \left[\sum_{i=1}^N \pi_t^i \|Y_t^i\|_2^2 - \frac{1}{2} \sum_{i=1}^N \pi_{t+L}^i \sum_{j=1}^N \sum_{l=1}^N w_{t:t+L-1}^{ij} w_{t:t+L-1}^{il} \|Y_t^j - Y_t^l\|_2^2 \right].
\end{aligned}$$

From Lemma 3, we have

$$\sum_{i=1}^N \pi_{t+L}^i \sum_{j=1}^N \sum_{k=1}^N w_{t:t+L-1}^{ij} w_{t:t+L-1}^{ik} \|Y_t^j - Y_t^k\|_2^2 \geq \frac{\pi_{\min}\beta^{2L}}{\delta_{\max}} \sum_{i=1}^N \pi_t^i \|Y_t^i\|_2^2,$$

which implies that

$$\begin{aligned}
&\sum_{i=1}^N \pi_{t+L}^i \| \left(\Pi_{k=t}^{t+L-1}(I + \alpha A(X_k))\right) \sum_{j=1}^N w_{t:t+L-1}^{ij} Y_t^j \|_2^2 \\
&\leq (1 + \alpha A_{\max})^{2L} (1 - \frac{\pi_{\min}\beta^{2L}}{2\delta_{\max}}) \sum_{i=1}^N \pi_t^i \|Y_t^i\|_2^2. && (33)
\end{aligned}$$

As for (31), since for any agent i we have $\|b^i(X_t) - B^\top(X_t)\pi_{t+1}\|_2 \le 2b_{\max}$ for all i, then

$$\begin{aligned}
&\|\hat{b}^i_{t+L}\|_2 \le \|(b^i(X_{t+L-1}) - B(X_{t+L-1})^\top \pi_{t+L})\|_2 \\
&\qquad + \sum_{k=t}^{t+L-2} \| \left(\Pi_{j=k+1}^{t+L-1}(I + \alpha A(X_j))\right) \|_2 \sum_{j=1}^{N} w^{ij}_{k+1:t+L-1} \|(b^j(X_k) - B(X_k)^\top \pi_{k+1})\|_2 \\
&\qquad \le 2b_{\max} \sum_{j=0}^{L-1} (1+\alpha A_{\max})^j \le 2b_{\max}(1+\alpha A_{\max})^{L-1} \sum_{j=0}^{L-1} \frac{1}{(1+\alpha A_{\max})^j} \\
&\qquad \le 2b_{\max} \frac{(1+\alpha A_{\max})^L - 1}{\alpha A_{\max}},
\end{aligned}$$

which implies that

$$\alpha^2 \sum_{i=1}^{N} \pi^i_{t+L} \|\hat{b}^i_{t+L}\|_2^2 \le \frac{4b^2_{\max}}{A^2_{\max}} \left((1+\alpha A_{\max})^L - 1\right)^2. \tag{34}$$

In addition, since for any vector x, there holds $2\|x\|_2 \le 1 + \|x\|_2^2$, then, for (32), we have

$$\begin{aligned}
&2\alpha \sum_{i=1}^{N} \pi^i_{t+L} (\hat{b}^i_{t+L})^\top \left(\Pi_{k=t}^{t+L-1}(I + \alpha A(X_k))\right) \sum_{j=1}^{N} w^{ij}_{t:t+L-1} Y^j_t \\
&\le 2\alpha \sum_{i=1}^{N} \pi^i_{t+L} \|\hat{b}^i_{t+L}\|_2 \|\Pi_{k=t}^{t+L-1}(I + \alpha A(X_k))\|_2 \sum_{j=1}^{N} w^{ij}_{t:t+L-1} \|Y^j_t\|_2 \\
&\le 4\alpha b_{\max} \frac{(1+\alpha A_{\max})^L - 1}{\alpha A_{\max}} (1+\alpha A_{\max})^L \sum_{i=1}^{N} \pi^i_t \|Y^i_t\|_2 \\
&\le 2b_{\max} \frac{(1+\alpha A_{\max})^L - 1}{A_{\max}} (1+\alpha A_{\max})^L \left(\sum_{i=1}^{N} \pi^i_t \|Y^i_t\|_2^2 + 1\right).
\end{aligned} \tag{35}$$

From (33)–(35), we have

$$\begin{aligned}
&\|Y_{t+L}\|^2_{M_{t+L}} \\
&\le (1+\alpha A_{\max})^{2L}(1 - \frac{\pi_{\min}\beta^{2L}}{2\delta_{\max}}) \sum_{i=1}^{N} \pi^i_t \|Y^i_t\|_2^2 + \frac{4b^2_{\max}}{A^2_{\max}} \left((1+\alpha A_{\max})^L - 1\right)^2 \\
&\quad + 2b_{\max} \frac{(1+\alpha A_{\max})^L - 1}{A_{\max}} (1+\alpha A_{\max})^L \left(\sum_{i=1}^{N} \pi^i_t \|Y^i_t\|_2^2 + 1\right) \\
&= \left((1+\alpha A_{\max})^{2L}(1 - \frac{\pi_{\min}\beta^{2L}}{2\delta_{\max}}) + 2b_{\max} \frac{(1+\alpha A_{\max})^L - 1}{A_{\max}} (1+\alpha A_{\max})^L\right) \|Y_t\|^2_{M_t} \\
&\quad + \frac{4b^2_{\max}}{A^2_{\max}} \left((1+\alpha A_{\max})^L - 1\right)^2 + 2b_{\max} \frac{(1+\alpha A_{\max})^L - 1}{A_{\max}} (1+\alpha A_{\max})^L.
\end{aligned}$$

From Lemma 2, $0 < \epsilon < 1$ when $0 < \alpha < \zeta_1$. With the definition of ϵ and ζ_2 in (6) and (12), we have

$$\begin{aligned}
\|Y_{t+L}\|^2_{M_{t+L}} &\le \epsilon \|Y_t\|^2_{M_t} + \zeta_2 \le \epsilon^{q_{t+L}} \|Y_{m_t}\|^2_{M_{m_t}} + \zeta_2 \sum_{k=0}^{q_{t+L}-1} \epsilon^k \\
&\le \epsilon^{q_{t+L}} \|Y_{m_t}\|^2_{M_{m_t}} + \frac{\zeta_2}{1-\epsilon},
\end{aligned}$$

which implies that

$$\sum_{i=1}^{N} \pi^i_t \|\theta^i_t - \langle\theta\rangle_t\|_2^2 \le \epsilon^{q_t} \sum_{i=1}^{N} \pi^i_{m_t} \|\theta^i_{m_t} - \langle\theta\rangle_{m_t}\|_2^2 + \frac{\zeta_2}{1-\epsilon},$$

where q_t and m_t are defined in Theorem 3. This completes the proof. ■

Lemma 5 *Suppose that Assumptions 2 and 3 hold. If $\{\mathbb{G}_t\}$ is uniformly strongly connected, then when the step-size α and corresponding mixing time $\tau(\alpha)$ satisfy*

$$0 < \alpha\tau(\alpha) < \frac{\log 2}{A_{\max}},$$

we have for any $t \geq \tau(\alpha)$,

$$\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2 \leq 2\alpha A_{\max}\tau(\alpha)\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 + 2\alpha\tau(\alpha)b_{\max} \tag{36}$$

$$\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2 \leq 6\alpha\tau(\alpha)A_{\max}\|\langle\theta\rangle_t\|_2 + 5\alpha\tau(\alpha)b_{\max} \tag{37}$$

$$\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 \leq 72\alpha^2\tau^2(\alpha)A_{\max}^2\|\langle\theta\rangle_t\|_2^2 + 50\alpha^2\tau^2(\alpha)b_{\max}^2 \leq 8\|\langle\theta\rangle_t\|_2^2 + \frac{6b_{\max}^2}{A_{\max}^2}. \tag{38}$$

Proof of Lemma 5: Recall the update of $\langle\theta\rangle_t$ at (4) with $\alpha_t = \alpha$ for all $t \geq 0$:

$$\langle\theta\rangle_{t+1} = \langle\theta\rangle_t + \alpha A(X_t)\langle\theta\rangle_t + \alpha B(X_t)^\top \pi_{t+1}.$$

Then, we have

$$\begin{aligned}\|\langle\theta\rangle_{t+1}\|_2 &\leq \|\langle\theta\rangle_t\|_2 + \alpha A_{\max}\|\langle\theta\rangle_t\|_2 + \alpha b_{\max}\\ &\leq (1+\alpha A_{\max})\|\langle\theta\rangle_t\|_2 + \alpha b_{\max}.\end{aligned}$$

By using $(1+x) \leq \exp(x)$, for all $u \in [t-\tau(\alpha), t]$, we have

$$\begin{aligned}\|\langle\theta\rangle_u\|_2 &\leq (1+\alpha A_{\max})^{u-t+\tau(\alpha)}\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 + \alpha b_{\max}\sum_{l=t-\tau(\alpha)}^{u-1}(1+\alpha A_{\max})^{u-1-l}\\ &\leq (1+\alpha A_{\max})^{\tau(\alpha)}\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 + \alpha b_{\max}\sum_{l=t-\tau(\alpha)}^{u-1}(1+\alpha A_{\max})^{u-1-t+\tau(\alpha)}\\ &\leq \exp(\alpha\tau(\alpha)A_{\max})\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 + \alpha\tau(\alpha)b_{\max}\exp(\alpha\tau(\alpha)A_{\max}).\end{aligned}$$

Since we have $\alpha\tau(\alpha)A_{\max} \leq \log 2 < \frac{1}{3}$, then $\exp(\alpha\tau(\alpha)A_{\max}) \leq 2$, which meas that

$$\|\langle\theta\rangle_u\|_2 \leq 2\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 + 2\alpha\tau(\alpha)b_{\max}.$$

Thus, we can use this to prove (36) for all $t \geq \tau(\alpha)$, i.e.,

$$\begin{aligned}\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2 &\leq \sum_{u=t-\tau(\alpha)}^{t-1}\|\langle\theta\rangle_{u+1} - \langle\theta\rangle_u\|_2\\ &\leq \alpha A_{\max}\sum_{u=t-\tau(\alpha)}^{t-1}\|\langle\theta\rangle_u\|_2 + \alpha\tau(\alpha)b_{\max}\\ &\leq \alpha A_{\max}\sum_{u=t-\tau(\alpha)}^{t-1}\big(2\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 + 2\alpha\tau(\alpha)b_{\max}\big) + \alpha\tau(\alpha)b_{\max}\\ &\leq 2\alpha\tau(\alpha)A_{\max}\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 + 2\alpha^2\tau^2(\alpha)A_{\max}b_{\max} + \alpha\tau(\alpha)b_{\max}\\ &\leq 2\alpha\tau(\alpha)A_{\max}\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 + \frac{5}{3}\alpha\tau(\alpha)b_{\max}\\ &\leq 2\alpha\tau(\alpha)A_{\max}\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 + 2\alpha\tau(\alpha)b_{\max}.\end{aligned}$$

Moreover, we can prove (37) by using the equation above for all $t \geq \tau(\alpha)$ as follows:

$$\begin{aligned}\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2 &\leq 2\alpha\tau(\alpha)A_{\max}\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 + \frac{5}{3}\alpha\tau(\alpha)b_{\max}\\ &\leq \frac{2}{3}\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2 + 2\alpha\tau(\alpha)A_{\max}\|\langle\theta\rangle_t\|_2 + \frac{5}{3}\alpha\tau(\alpha)b_{\max}\\ &\leq 6\alpha\tau(\alpha)A_{\max}\|\langle\theta\rangle_t\|_2 + 5\alpha\tau(\alpha)b_{\max}.\end{aligned}$$

Next, using the inequality $(x+y)^2 \leq 2x^2 + y^2$ for all x, y, we can show (38) with (37), i.e.,

$$\begin{aligned}\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 &\leq 72\alpha^2\tau^2(\alpha)A_{\max}^2\|\langle\theta\rangle_t\|_2^2 + 50\alpha^2\tau^2(\alpha)b_{\max}^2 \\ &\leq 8\|\langle\theta\rangle_t\|_2^2 + \frac{6b_{\max}^2}{A_{\max}^2},\end{aligned}$$

where we use $\alpha\tau(\alpha)A_{\max} < \frac{1}{3}$ in the last inequality. ■

Lemma 6 *Let $\mathcal{F}_t = \sigma(X_k,\ k \leq t)$ be a σ-algebra on $\{X_t\}$. Suppose that Assumptions 2–4 and 6 hold. If $\{\mathbb{G}_t\}$ is uniformly strongly connected, then when*

$$0 < \alpha < \frac{\log 2}{A_{\max}\tau(\alpha)},$$

we have for any $t \geq \tau(\alpha)$,

$$\begin{aligned}&\left|\mathbf{E}\left[(\langle\theta\rangle_t - \theta^*)^\top(P+P^\top)\left(A(X_t)\langle\theta\rangle_t + B(X_t)^\top\pi_{t+1} - A\langle\theta\rangle_t - b\right) \mid \mathcal{F}_{t-\tau(\alpha)}\right]\right| \\ &\leq \alpha\gamma_{\max}\left(72 + 456\tau(\alpha)A_{\max}^2 + 84\tau(\alpha)A_{\max}b_{\max}\right)\mathbf{E}\left[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}\right] \\ &\quad + \alpha\gamma_{\max}\left[2 + 4\|\theta^*\|_2^2 + \frac{48b_{\max}^2}{A_{\max}^2} + \tau(\alpha)A_{\max}^2\left(152\left(\frac{b_{\max}}{A_{\max}} + \|\theta^*\|_2\right)^2\right.\right. \\ &\quad \left.\left. + \frac{48b_{\max}}{A_{\max}}\left(\frac{b_{\max}}{A_{\max}} + 1\right)^2 + \frac{87b_{\max}^2}{A_{\max}^2} + \frac{12b_{\max}}{A_{\max}}\right)\right] \\ &\quad + 2\gamma_{\max}\eta_{t+1}\sqrt{N}b_{\max}\left(1 + 9\mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + \frac{6b_{\max}^2}{A_{\max}^2} + \|\theta^*\|_2^2\right).\end{aligned}$$

Proof of Lemma 6: Note that for all $t \geq \tau(\alpha)$, we have

$$\begin{aligned}&|\mathbf{E}[(\langle\theta\rangle_t - \theta^*)^\top(P+P^\top)(A(X_t)\langle\theta\rangle_t + B(X_t)^\top\pi_{t+1} - A\langle\theta\rangle_t - b) \mid \mathcal{F}_{t-\tau(\alpha)}]| \\ &\leq |\mathbf{E}[(\langle\theta\rangle_t - \theta^*)^\top(P+P^\top)(A(X_t) - A)\langle\theta\rangle_t \mid \mathcal{F}_{t-\tau(\alpha)}]| \\ &\quad + |\mathbf{E}[(\langle\theta\rangle_t - \theta^*)^\top(P+P^\top)(B(X_t)^\top\pi_{t+1} - b) \mid \mathcal{F}_{t-\tau(\alpha)}]| \\ &\leq |\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*)^\top(P+P^\top)(A(X_t) - A)\langle\theta\rangle_{t-\tau(\alpha)} \mid \mathcal{F}_{t-\tau(\alpha)}]| \quad (39) \\ &\quad + |\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*)^\top(P+P^\top)(A(X_t) - A)(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}) \mid \mathcal{F}_{t-\tau(\alpha)}]| \quad (40) \\ &\quad + |\mathbf{E}[(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)})^\top(P+P^\top)(A(X_t) - A)\langle\theta\rangle_{t-\tau(\alpha)} \mid \mathcal{F}_{t-\tau(\alpha)}]| \quad (41) \\ &\quad + |\mathbf{E}[(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)})^\top(P+P^\top)(A(X_t) - A)(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}) \mid \mathcal{F}_{t-\tau(\alpha)}]| \quad (42) \\ &\quad + |\mathbf{E}[(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)})^\top(P+P^\top)(B(X_t)^\top\pi_{t+1} - b) \mid \mathcal{F}_{t-\tau(\alpha)}]| \quad (43) \\ &\quad + |\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*)^\top(P+P^\top)(B(X_t)^\top\pi_{t+1} - b) \mid \mathcal{F}_{t-\tau(\alpha)}]|. \quad (44)\end{aligned}$$

First, by using the mixing time in Assumption 3, we can get the bound for (39) and (44) for all $t \geq \tau(\alpha)$ as follows:

$$\begin{aligned}&|\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*)^\top(P+P^\top)(A(X_t) - A)\langle\theta\rangle_{t-\tau(\alpha)} \mid \mathcal{F}_{t-\tau(\alpha)}]| \\ &\leq |(\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*)^\top(P+P^\top)\mathbf{E}[A(X_t) - A \mid \mathcal{F}_{t-\tau(\alpha)}]\langle\theta\rangle_{t-\tau(\alpha)}| \\ &\leq 2\alpha\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*\|_2\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\ &\leq \alpha\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*\|_2^2 + \|\langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\ &\leq \alpha\gamma_{\max}\mathbf{E}[2\|\theta^*\|_2^2 + 3\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\ &\leq 6\alpha\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 6\alpha\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 2\alpha\gamma_{\max}\|\theta^*\|_2^2 \\ &\leq 54\alpha\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 36\alpha\gamma_{\max}(\frac{b_{\max}}{A_{\max}})^2 + 2\alpha\gamma_{\max}\|\theta^*\|_2^2, \quad (45)\end{aligned}$$

where in the last inequality, we use (36) from Lemma 5. Then, from the definition of π_∞ in Assumption 6,

$$
\begin{aligned}
&|\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*)^\top (P + P^\top)(B(X_t)^\top \pi_{t+1} - b) \mid \mathcal{F}_{t-\tau(\alpha)}]| \\
&\le |\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*)^\top (P + P^\top)(\sum_{i=1}^N \pi_{t+1}^i (b^i(X_t) - b^i) + \sum_{i=1}^N (\pi_{t+1}^i - \pi_\infty^i) b^i) \mid \mathcal{F}_{t-\tau(\alpha)}]| \\
&\le |(\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*)^\top (P + P^\top)(\sum_{i=1}^N \pi_{t+1}^i \mathbf{E}[b^i(X_t) - b^i \mid \mathcal{F}_{t-\tau(\alpha)}] + \sum_{i=1}^N (\pi_{t+1}^i - \pi_\infty^i) b^i)| \\
&\le 2\gamma_{\max}(\alpha + \eta_{t+1}\sqrt{N} b_{\max}) \mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*\|_2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\
&\le 2\gamma_{\max}(\alpha + \eta_{t+1}\sqrt{N} b_{\max}) \left(\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha)}] + \|\theta^*\|_2\right) \\
&\le 2\gamma_{\max}(\alpha + \eta_{t+1}\sqrt{N} b_{\max}) \left(1 + \frac{1}{2}\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + \frac{1}{2}\|\theta^*\|_2^2\right) \\
&\le 2\gamma_{\max}(\alpha + \eta_{t+1}\sqrt{N} b_{\max}) \left(1 + \mathbf{E}[\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 + \|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + \|\theta^*\|_2^2\right) \\
&\le 2\gamma_{\max}(\alpha + \eta_{t+1}\sqrt{N} b_{\max}) \left(1 + 9\mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 6(\frac{b_{\max}}{A_{\max}})^2 + \|\theta^*\|_2^2\right),
\end{aligned}
\tag{46}
$$

where we also use (36) from Lemma 5 in the last inequality.

Next, by using Assumption 2, (36) and (38), we have

$$
\begin{aligned}
&|\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*)^\top (P + P^\top)(A(X_t) - A)(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}) \mid \mathcal{F}_{t-\tau(\alpha)}]| \\
&\le 4\gamma_{\max} A_{\max} \mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)} - \theta^*\|_2 \|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\
&\le 4\gamma_{\max} A_{\max} \mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 \|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\
&\quad + 4\gamma_{\max} A_{\max} \|\theta^*\|_2 \mathbf{E}[\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\
&\le 8\alpha\tau(\alpha)\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 8\alpha\tau(\alpha)\gamma_{\max} A_{\max} b_{\max} \|\theta^*\|_2 \\
&\quad + 8\alpha\tau(\alpha)\gamma_{\max} A_{\max}^2 \left(\frac{b_{\max}}{A_{\max}} + \|\theta^*\|_2\right) \mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\
&\le 8\alpha\tau(\alpha)\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 8\alpha\tau(\alpha)\gamma_{\max} A_{\max} b_{\max} \|\theta^*\|_2 \\
&\quad + 4\alpha\tau(\alpha)\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 4\alpha\tau(\alpha)\gamma_{\max} A_{\max}^2 \left(\frac{b_{\max}}{A_{\max}} + \|\theta^*\|_2\right)^2 \\
&\le 12\alpha\tau(\alpha)\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 8\alpha\tau(\alpha)\gamma_{\max} \left(b_{\max} + A_{\max}\|\theta^*\|_2\right)^2 \\
&\le 24\alpha\tau(\alpha)\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 8\alpha\tau(\alpha)\gamma_{\max} \left(b_{\max} + A_{\max}\|\theta^*\|_2\right)^2 \\
&\quad + 24\alpha\tau(\alpha)\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\
&\le 216\alpha\tau(\alpha)\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 144\alpha\tau(\alpha)\gamma_{\max} b_{\max}^2 \\
&\quad + 8\alpha\tau(\alpha)\gamma_{\max} \left(b_{\max} + A_{\max}\|\theta^*\|_2\right)^2 \\
&\le 216\alpha\tau(\alpha)\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 152\alpha\tau(\alpha)\gamma_{\max} \left(b_{\max} + A_{\max}\|\theta^*\|_2\right)^2.
\end{aligned}
\tag{47}
$$

In additional, by using (36) and (38), we have

$$
\begin{aligned}
&|\mathbf{E}[(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)})^\top (P + P^\top)(A(X_t) - A)\langle\theta\rangle_{t-\tau(\alpha)} \mid \mathcal{F}_{t-\tau(\alpha)}]| \\
&\le 4\gamma_{\max} A_{\max} \mathbf{E}[\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2 \|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\
&\le 8\alpha\tau(\alpha)\gamma_{\max} A_{\max} \mathbf{E}[A_{\max}\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 + b_{\max}\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\
&\le 4\alpha\tau(\alpha)\gamma_{\max} A_{\max}(2A_{\max} + b_{\max}) \mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 4\alpha\tau(\alpha)\gamma_{\max} A_{\max} b_{\max} \\
&\le 8\alpha\tau(\alpha)\gamma_{\max} A_{\max}(2A_{\max} + b_{\max}) \mathbf{E}[\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\
&\quad + 8\alpha\tau(\alpha)\gamma_{\max} A_{\max}(2A_{\max} + b_{\max}) \mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] + 4\alpha\tau(\alpha)\gamma_{\max} A_{\max} b_{\max} \\
&\le 72\alpha\tau(\alpha)\gamma_{\max} A_{\max}(2A_{\max} + b_{\max}) \mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}] \\
&\quad + 48\alpha\tau(\alpha)\gamma_{\max} A_{\max} b_{\max} (\frac{b_{\max}}{A_{\max}} + 1)^2.
\end{aligned}
\tag{48}
$$

Moreover, we can get the bound for (42) by using (38) as follows:

$$\begin{aligned}
&|\mathbf{E}[(\langle\theta\rangle_t-\langle\theta\rangle_{t-\tau(\alpha)})^\top(P+P^\top)(A(X_t)-A)(\langle\theta\rangle_t-\langle\theta\rangle_{t-\tau(\alpha)})\mid\mathcal{F}_{t-\tau(\alpha)}]|\\
&\le 4\gamma_{\max}A_{\max}\mathbf{E}[\|\langle\theta\rangle_t-\langle\theta\rangle_{t-\tau(\alpha)}\|_2^2\mid\mathcal{F}_{t-\tau(\alpha)}]|\\
&\le 4\gamma_{\max}A_{\max}\mathbf{E}[72\alpha^2\tau^2(\alpha)A_{\max}^2\|\langle\theta\rangle_t\|_2^2+50\alpha^2\tau^2(\alpha)b_{\max}^2\mid\mathcal{F}_{t-\tau(\alpha)}]\\
&\le 96\alpha\tau(\alpha)A_{\max}^2\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha)}]+67\alpha\tau(\alpha)b_{\max}^2\gamma_{\max}.
\end{aligned}\tag{49}$$

Finally, using (37) we can get the bound for (43):

$$\begin{aligned}
&|\mathbf{E}[(\langle\theta\rangle_t-\langle\theta\rangle_{t-\tau(\alpha)})^\top(P+P^\top)(B(X_t)^\top\pi_{t+1}-b)\mid\mathcal{F}_{t-\tau(\alpha)}]|\\
&\le 4\gamma_{\max}b_{\max}\mathbf{E}[\|\langle\theta\rangle_t-\langle\theta\rangle_{t-\tau(\alpha)}\|_2\mid\mathcal{F}_{t-\tau(\alpha)}]\\
&\le 4\gamma_{\max}b_{\max}\mathbf{E}[6\alpha\tau(\alpha)A_{\max}\|\langle\theta\rangle_t\|_2+5\alpha\tau(\alpha)b_{\max}\mid\mathcal{F}_{t-\tau(\alpha)}]\\
&\le 12\alpha\tau(\alpha)\gamma_{\max}A_{\max}b_{\max}\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha)}]+12\alpha\tau(\alpha)\gamma_{\max}A_{\max}b_{\max}+20\alpha\tau(\alpha)b_{\max}^2\gamma_{\max}.
\end{aligned}\tag{50}$$

Then, by using (45)–(50), we have

$$\begin{aligned}
&|\mathbf{E}[(\langle\theta\rangle_t^\top-\theta^*)^\top(P+P^\top)(A(X_t)\langle\theta\rangle_t+B(X_t)^\top\pi_{t+1}-A\langle\theta\rangle_t-b)\mid\mathcal{F}_{t-\tau(\alpha)}]|\\
&\le 54\alpha\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha)}]+36\alpha\gamma_{\max}(\frac{b_{\max}}{A_{\max}})^2+2\alpha\gamma_{\max}\|\theta^*\|_2^2\\
&\quad+216\alpha\tau(\alpha)\gamma_{\max}A_{\max}^2\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha)}]+152\alpha\tau(\alpha)\gamma_{\max}\left(b_{\max}+A_{\max}\|\theta^*\|_2\right)^2\\
&\quad+72\alpha\tau(\alpha)\gamma_{\max}A_{\max}(2A_{\max}+b_{\max})\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha)}]\\
&\quad+48\alpha\tau(\alpha)\gamma_{\max}A_{\max}b_{\max}(\frac{b_{\max}}{A_{\max}}+1)^2+20\alpha\tau(\alpha)b_{\max}^2\gamma_{\max}\\
&\quad+96\alpha\tau(\alpha)A_{\max}^2\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha)}]+67\alpha\tau(\alpha)b_{\max}^2\gamma_{\max}\\
&\quad+12\alpha\tau(\alpha)\gamma_{\max}A_{\max}b_{\max}\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha)}]+12\alpha\tau(\alpha)\gamma_{\max}A_{\max}b_{\max}\\
&\quad+2\gamma_{\max}(\alpha+\eta_{t+1}\sqrt{N}b_{\max})\left(1+9\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha)}]+6(\frac{b_{\max}}{A_{\max}})^2+\|\theta^*\|_2^2\right)\\
&\le\alpha\gamma_{\max}\left(72+456\tau(\alpha)A_{\max}^2+84\tau(\alpha)A_{\max}b_{\max}\right)\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha)}]\\
&\quad+\alpha\gamma_{\max}\Big[2+4\|\theta^*\|_2^2+48(\frac{b_{\max}}{A_{\max}})^2+\tau(\alpha)A_{\max}^2\Big(152\left(\frac{b_{\max}}{A_{\max}}+\|\theta^*\|_2\right)^2\\
&\quad+48\frac{b_{\max}}{A_{\max}}(\frac{b_{\max}}{A_{\max}}+1)^2+87(\frac{b_{\max}}{A_{\max}})^2+12\frac{b_{\max}}{A_{\max}}\Big)\Big]\\
&\quad+2\gamma_{\max}\eta_{t+1}\sqrt{N}b_{\max}\left(1+9\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha)}]+6(\frac{b_{\max}}{A_{\max}})^2+\|\theta^*\|_2^2\right).
\end{aligned}\tag{51}$$

This completes the proof. ■

Lemma 7 *Suppose that Assumptions 2–4 and 6 hold. Then, when*

$$0<\alpha<\min\left\{\frac{\log 2}{A_{\max}\tau(\alpha)},\ \frac{0.1}{K_2\gamma_{\max}}\right\},$$

we have for any $t\ge T_1$,

$$\begin{aligned}
\mathbf{E}\left[\|\langle\theta\rangle_t-\theta^*\|_2^2\right]&\le\left(1-\frac{0.9\alpha}{\gamma_{\max}}\right)^{t-T_1}\frac{\gamma_{\max}}{\gamma_{\min}}\mathbf{E}\left[\|\langle\theta\rangle_{T_1}-\theta^*\|_2^2\right]+\frac{\gamma_{\max}}{\gamma_{\min}}\cdot\frac{\alpha\zeta_3\gamma_{\max}+2\gamma_{\max}\zeta_4}{0.9}\\
&\le\left(1-\frac{0.9\alpha}{\gamma_{\max}}\right)^{t-T_1}\frac{\gamma_{\max}}{\gamma_{\min}}\left(4\exp\left\{2\alpha A_{\max}T_1\right\}+2\right)\mathbf{E}\left[\|\langle\theta\rangle_0-\theta^*\|_2^2\right]\\
&\quad+4\left(1-\frac{0.9\alpha}{\gamma_{\max}}\right)^{t-T_1}\frac{\gamma_{\max}}{\gamma_{\min}}\exp\left\{2\alpha A_{\max}T_1\right\}\left(\|\theta^*\|_2+\frac{b_{\max}}{A_{\max}}\right)^2\\
&\quad+\frac{\gamma_{\max}}{\gamma_{\min}}\cdot\frac{\alpha\zeta_3\gamma_{\max}+2\gamma_{\max}\zeta_4}{0.9},
\end{aligned}$$

where ζ_3, ζ_4 and K_2 are defined in(13), (14) *and* (20)*, respectively.*

Proof of Lemma 7: Let $H(\langle\theta\rangle_t) = (\langle\theta\rangle_t - \theta^*)^\top P(\langle\theta\rangle_t - \theta^*)$. From Assumption 4, we know that

$$\gamma_{\min}\|\langle\theta\rangle_t - \theta^*\|_2^2 \leq H(\langle\theta\rangle_t) \leq \gamma_{\max}\|\langle\theta\rangle_t - \theta^*\|_2^2.$$

Moreover, from Assumption 2, for all $t \geq 0$ we have

$$\begin{aligned}
&H(\langle\theta\rangle_{t+1}) \\
&= (\langle\theta\rangle_{t+1} - \theta^*)^\top P(\langle\theta\rangle_{t+1} - \theta^*) \\
&= (\langle\theta\rangle_t + \alpha A(X_t)\langle\theta\rangle_t + \alpha B(X_t)^\top \pi_{t+1} - \theta^*)^\top P(\langle\theta\rangle_t + \alpha A(X_t)\langle\theta\rangle_t + \alpha B(X_t)^\top \pi_{t+1} - \theta^*) \\
&= (\langle\theta\rangle_t - \theta^*)^\top P(\langle\theta\rangle_t - \theta^*) + \alpha^2 (A(X_t)\langle\theta\rangle_t)^\top P(A(X_t)\langle\theta\rangle_t) \\
&\quad + \alpha^2 (B(X_t)^\top \pi_{t+1})^\top P(B(X_t)^\top \pi_{t+1}) + \alpha^2 (A(X_t)\langle\theta\rangle_t)^\top (P + P^\top)(B(X_t)^\top \pi_{t+1}) \\
&\quad + \alpha(\langle\theta\rangle_t - \theta^*)^\top (P + P^\top)(A(X_t)\langle\theta\rangle_t + B(X_t)^\top \pi_{t+1} - A\langle\theta\rangle_t - b) \\
&\quad + \alpha(\langle\theta\rangle_t - \theta^*)^\top P(A\langle\theta\rangle_t + b) + \alpha(A\langle\theta\rangle_t + b)^\top P(\langle\theta\rangle_t - \theta^*) \\
&= H(\langle\theta\rangle_t) + \alpha^2 (A(X_t)\langle\theta\rangle_t)^\top P(A(X_t)\langle\theta\rangle_t) \\
&\quad + \alpha^2 (B(X_t)^\top \pi_{t+1})^\top P(B(X_t)^\top \pi_{t+1}) + \alpha^2 (A(X_t)\langle\theta\rangle_t)^\top (P + P^\top)(B(X_t)^\top \pi_{t+1}) \\
&\quad + \alpha(\langle\theta\rangle_t - \theta^*)^\top (P + P^\top)(A(X_t)\langle\theta\rangle_t + B(X_t)^\top \pi_{t+1} - A\langle\theta\rangle_t - b) \\
&\quad + \alpha(\langle\theta\rangle_t - \theta^*)^\top (PA + A^\top P)(\langle\theta\rangle_t - \theta^*), \qquad (52)
\end{aligned}$$

where we use the fact that $A\theta^* + b = 0$ on the last equality.

Next, we can take expectation on both sides of (52). From Assumption 4 and Lemma 6, for $t \geq T_1$ we have

$$\begin{aligned}
&\mathbf{E}[H(\langle\theta\rangle_{t+1})] \\
&= \mathbf{E}[H(\langle\theta\rangle_t)] + \alpha^2 \mathbf{E}[(A(X_t)\langle\theta\rangle_t)^\top P(A(X_t)\langle\theta\rangle_t)] - \alpha \mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] \\
&\quad + \alpha^2 \mathbf{E}[(B(X_t)^\top \pi_{t+1})^\top P(B(X_t)^\top \pi_{t+1})] + \alpha^2 \mathbf{E}[(A(X_t)\langle\theta\rangle_t)^\top (P + P^\top)(B(X_t)^\top \pi_{t+1})] \\
&\quad + \alpha \mathbf{E}[(\langle\theta\rangle_t - \theta^*)^\top (P + P^\top)(A(X_t)\langle\theta\rangle_t + B(X_t)^\top \pi_{t+1} - A\langle\theta\rangle_t - b)] \\
&\leq \mathbf{E}[H(\langle\theta\rangle_t)] - \alpha \mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] + \alpha^2 A_{\max}^2 \gamma_{\max} \mathbf{E}[\|\langle\theta\rangle_t\|_2^2] + 2\alpha^2 A_{\max} b_{\max} \gamma_{\max} \mathbf{E}[\|\langle\theta\rangle_t\|_2] \\
&\quad + \alpha^2 b_{\max}^2 \gamma_{\max} + \alpha^2 \gamma_{\max} \left(72 + 456\tau(\alpha) A_{\max}^2 + 84\tau(\alpha) A_{\max} b_{\max}\right) \mathbf{E}[\|\langle\theta\rangle_t\|_2^2] \\
&\quad + \alpha^2 \gamma_{\max} \Big[2 + 4\|\theta^*\|_2^2 + 48(\frac{b_{\max}}{A_{\max}})^2 + \tau(\alpha) A_{\max}^2 \Big(152 \left(\frac{b_{\max}}{A_{\max}} + \|\theta^*\|_2\right)^2 \\
&\quad + 48\frac{b_{\max}}{A_{\max}}(\frac{b_{\max}}{A_{\max}} + 1)^2 + 87(\frac{b_{\max}}{A_{\max}})^2 + 12\frac{b_{\max}}{A_{\max}}\Big)\Big] \\
&\quad + 2\alpha \gamma_{\max} \eta_{t+1} \sqrt{N} b_{\max} \left(1 + 9\mathbf{E}[\|\langle\theta\rangle_t\|_2^2] + 6(\frac{b_{\max}}{A_{\max}})^2 + \|\theta^*\|_2^2\right) \\
&\leq \mathbf{E}[H(\langle\theta\rangle_t)] - \alpha \mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] + 2\alpha^2 A_{\max}^2 \gamma_{\max} \mathbf{E}[\|\langle\theta\rangle_t\|_2^2] + 2\alpha^2 b_{\max}^2 \gamma_{\max} \\
&\quad + \alpha^2 \gamma_{\max} \left(72 + 456\tau(\alpha) A_{\max}^2 + 84\tau(\alpha) A_{\max} b_{\max}\right) \mathbf{E}[\|\langle\theta\rangle_t\|_2^2] \\
&\quad + \alpha^2 \gamma_{\max} \Big[2 + 4\|\theta^*\|_2^2 + 48(\frac{b_{\max}}{A_{\max}})^2 + \tau(\alpha) A_{\max}^2 \Big(152 \left(\frac{b_{\max}}{A_{\max}} + \|\theta^*\|_2\right)^2 \\
&\quad + 48\frac{b_{\max}}{A_{\max}}(\frac{b_{\max}}{A_{\max}} + 1)^2 + 87(\frac{b_{\max}}{A_{\max}})^2 + 12\frac{b_{\max}}{A_{\max}}\Big)\Big] \\
&\quad + 2\alpha \gamma_{\max} \eta_{t+1} \sqrt{N} b_{\max} \left(1 + 9\mathbf{E}[\|\langle\theta\rangle_t\|_2] + 6(\frac{b_{\max}}{A_{\max}})^2 + \|\theta^*\|_2^2\right).
\end{aligned}$$

743 Since $\mathbf{E}[\|\langle\theta\rangle_t\|_2^2] \leq 2\mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] + 2\|\theta^*\|_2^2$, we have

$$\begin{aligned}
&\mathbf{E}[H(\langle\theta\rangle_{t+1})] \\
&\leq \mathbf{E}[H(\langle\theta\rangle_t)] - \alpha\mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] + 2\alpha^2 b_{\max}^2\gamma_{\max} \\
&\quad + \alpha^2\gamma_{\max}\left(72 + 2A_{\max}^2 + 456\tau(\alpha)A_{\max}^2 + 84\tau(\alpha)A_{\max}b_{\max}\right)(2\mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] + 2\|\theta^*\|_2^2) \\
&\quad + \alpha^2\gamma_{\max}\Big[2 + 4\|\theta^*\|_2^2 + 48(\frac{b_{\max}}{A_{\max}})^2 + \tau(\alpha)A_{\max}^2\Big(152\left(\frac{b_{\max}}{A_{\max}} + \|\theta^*\|_2\right)^2 \\
&\quad + 48\frac{b_{\max}}{A_{\max}}(\frac{b_{\max}}{A_{\max}} + 1)^2 + 87(\frac{b_{\max}}{A_{\max}})^2 + 12\frac{b_{\max}}{A_{\max}}\Big)\Big] \\
&\quad + 2\alpha\gamma_{\max}\eta_{t+1}\sqrt{N}b_{\max}\left(1 + 18\mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] + 6(\frac{b_{\max}}{A_{\max}})^2 + 19\|\theta^*\|_2^2\right) \\
&\leq \mathbf{E}[H(\langle\theta\rangle_t)] \\
&\quad + \left(-\alpha + 2\alpha^2\gamma_{\max}\left(72 + 2A_{\max}^2 + 456\tau(\alpha)A_{\max}^2 + 84\tau(\alpha)A_{\max}b_{\max}\right)\right)\mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] \\
&\quad + 2\alpha^2\gamma_{\max}\left(72 + 2A_{\max}^2 + 456\tau(\alpha)A_{\max}^2 + 84\tau(\alpha)A_{\max}b_{\max}\right)\|\theta^*\|_2^2 \\
&\quad + \alpha^2\gamma_{\max}\Big[2 + 2b_{\max}^2 + 4\|\theta^*\|_2^2 + 48(\frac{b_{\max}}{A_{\max}})^2 + \tau(\alpha)A_{\max}^2\Big(152\left(\frac{b_{\max}}{A_{\max}} + \|\theta^*\|_2\right)^2 \\
&\quad + 48\frac{b_{\max}}{A_{\max}}(\frac{b_{\max}}{A_{\max}} + 1)^2 + 87(\frac{b_{\max}}{A_{\max}})^2 + 12\frac{b_{\max}}{A_{\max}}\Big)\Big] \\
&\quad + 2\alpha\gamma_{\max}\eta_{t+1}\sqrt{N}b_{\max}\left(1 + 18\mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] + 6(\frac{b_{\max}}{A_{\max}})^2 + 19\|\theta^*\|_2^2\right) \\
&\leq \mathbf{E}[H(\langle\theta\rangle_t)] + \left(-\alpha + \alpha^2\gamma_{\max}K_2 + 36\alpha\eta_{t+1}\sqrt{N}b_{\max}\gamma_{\max}\right)\mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] \\
&\quad + \alpha^2\zeta_3\gamma_{\max} + \alpha\gamma_{\max}\eta_{t+1}\zeta_4.
\end{aligned}$$

744 From Lemma 2, $1 - \frac{0.9\alpha}{\gamma_{\max}} \in (0, 1)$. In addition, from the definition of T_1 and $\alpha < \frac{0.1}{K_2\gamma_{\max}}$, we have

$$\begin{aligned}
&\mathbf{E}[H(\langle\theta\rangle_{t+1})] \\
&\leq \mathbf{E}[H(\langle\theta\rangle_t)] - 0.9\alpha\mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] + \alpha^2\zeta_3\gamma_{\max} + \alpha\gamma_{\max}\eta_{t+1}\zeta_4 \\
&\leq \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)\mathbf{E}[H(\langle\theta\rangle_t)] + \alpha^2\zeta_3\gamma_{\max} + \alpha\gamma_{\max}\eta_{t+1}\zeta_4 \\
&\leq \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{t+1-T_1}\mathbf{E}[H(\langle\theta\rangle_{T_1})] + \left(\alpha^2\zeta_3\gamma_{\max} + 2\alpha\gamma_{\max}\zeta_4\right)\sum_{k=T_1}^{t}(1 - \frac{0.9\alpha}{\gamma_{\max}})^{t-k} \\
&\leq \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{t+1-T_1}\mathbf{E}[H(\langle\theta\rangle_{T_1})] + (\alpha\zeta_3\gamma_{\max} + 2\gamma_{\max}\zeta_4)\frac{\gamma_{\max}}{0.9},
\end{aligned} \tag{53}$$

745 which implies that

$$\begin{aligned}
&\mathbf{E}[\|\langle\theta\rangle_{t+1} - \theta^*\|_2^2] \\
&\leq \frac{1}{\gamma_{\min}}\mathbf{E}[H(\langle\theta\rangle_{t+1})] \\
&\leq \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{t+1-T_1}\frac{\gamma_{\max}}{\gamma_{\min}}\mathbf{E}[\|\langle\theta\rangle_{T_1} - \theta^*\|_2^2] + \frac{\gamma_{\max}}{\gamma_{\min}}\frac{\alpha\zeta_3\gamma_{\max} + 2\gamma_{\max}\zeta_4}{0.9}.
\end{aligned} \tag{54}$$

Next, we consider the bound for $\mathbf{E}[\|\langle\theta\rangle_{T_1} - \theta^*\|_2^2]$. Since $1 + x \leq \exp\{x\}$ for any x, we have for any t,

$$\begin{aligned}
&\|\langle\theta\rangle_{t+1} - \langle\theta\rangle_0\|_2 \\
&= \|\langle\theta\rangle_t - \langle\theta\rangle_0 + \alpha A(X_t)(\langle\theta\rangle_t - \langle\theta\rangle_0) + \alpha B(X_t)^\top \pi_{t+1} + \alpha A(X_t)\langle\theta\rangle_0\|_2 \\
&\leq (1 + \alpha A_{\max})\|\langle\theta\rangle_t - \langle\theta\rangle_0\|_2 + \alpha\left(A_{\max}\|\langle\theta\rangle_0\|_2 + b_{\max}\right) \\
&\leq \alpha\left(A_{\max}\|\langle\theta\rangle_0\|_2 + b_{\max}\right)\sum_{l=0}^{t}(1 + \alpha A_{\max})^l \\
&\leq \left(A_{\max}\|\langle\theta\rangle_0\|_2 + b_{\max}\right)\frac{(1 + \alpha A_{\max})^{t+1}}{A_{\max}} \\
&\leq \left(\|\langle\theta\rangle_0 - \theta^*\|_2 + \|\theta^*\|_2 + \frac{b_{\max}}{A_{\max}}\right)\exp\left\{\alpha A_{\max}(t+1)\right\},
\end{aligned}$$

which implies that

$$\|\langle\theta\rangle_{T_1} - \langle\theta\rangle_0\|_2 \leq \left(\|\langle\theta\rangle_0 - \theta^*\|_2 + \|\theta^*\|_2 + \frac{b_{\max}}{A_{\max}}\right)\exp\left\{\alpha A_{\max} T_1\right\}.$$

Then, we have

$$\begin{aligned}
\mathbf{E}[\|\langle\theta\rangle_{T_1} - \theta^*\|_2^2] &\leq 2\|\langle\theta\rangle_{T_1} - \langle\theta\rangle_0\|_2^2 + 2\|\langle\theta\rangle_0 - \theta^*\|_2^2 \\
&\leq (4\exp\left\{2\alpha A_{\max} T_1\right\} + 2)\mathbf{E}[\|\langle\theta\rangle_0 - \theta^*\|_2^2] \\
&\quad + 4\exp\left\{2\alpha A_{\max} T_1\right\}(\|\theta^*\|_2 + \frac{b_{\max}}{A_{\max}})^2. \qquad (55)
\end{aligned}$$

From (54) and (55), we have

$$\begin{aligned}
\mathbf{E}[\|\langle\theta\rangle_{t+1} - \theta^*\|_2^2] \leq &\left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{t+1-T_1}\frac{\gamma_{\max}}{\gamma_{\min}}(4\exp\left\{2\alpha A_{\max} T_1\right\} + 2)\mathbf{E}[\|\langle\theta\rangle_0 - \theta^*\|_2^2] \\
&+ 4\left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{t+1-T_1}\frac{\gamma_{\max}}{\gamma_{\min}}\exp\left\{2\alpha A_{\max} T_1\right\}(\|\theta^*\|_2 + \frac{b_{\max}}{A_{\max}})^2 \\
&+ \frac{\gamma_{\max}}{\gamma_{\min}}\frac{\alpha\zeta_3\gamma_{\max} + 2\gamma_{\max}\zeta_4}{0.9}.
\end{aligned}$$

This completes the proof. ■

We are now in a position to prove the fixed step-size case in Theorem 3.

Proof of Case 1) in Theorem 3: From Lemmas 4 and 7, for any $t \geq T_1$, we have

$$\begin{aligned}
&\sum_{i=1}^{N}\pi_t^i\mathbf{E}[\|\theta_t^i - \theta^*\|_2^2] \\
&\leq 2\sum_{i=1}^{N}\pi_t^i\mathbf{E}[\|\theta_t^i - \langle\theta\rangle_t\|_2^2] + 2\mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] \\
&\leq 2\epsilon^{q_t}\sum_{i=1}^{N}\pi_{m_t}^i\mathbf{E}[\|\theta_{m_t}^i - \langle\theta\rangle_{m_t}\|_2^2] + \frac{2\zeta_2}{1-\epsilon} + \frac{\gamma_{\max}}{\gamma_{\min}}\cdot\frac{2\alpha\zeta_3\gamma_{\max} + 4\gamma_{\max}\zeta_4}{0.9} \\
&\quad + \left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{t-T_1}\frac{\gamma_{\max}}{\gamma_{\min}}(8\exp\left\{2\alpha A_{\max} T_1\right\} + 4)\mathbf{E}[\|\langle\theta\rangle_0 - \theta^*\|_2^2] \\
&\quad + 8\left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{t-T_1}\frac{\gamma_{\max}}{\gamma_{\min}}\exp\left\{2\alpha A_{\max} T_1\right\}(\|\theta^*\|_2 + \frac{b_{\max}}{A_{\max}})^2 \\
&\leq 2\epsilon^{q_t}\sum_{i=1}^{N}\pi_{m_t}^i\mathbf{E}\left[\left\|\theta_{m_t}^i - \langle\theta\rangle_{m_t}\right\|_2^2\right] + C_1\left(1 - \frac{0.9\alpha}{\gamma_{\max}}\right)^{t-T_1} + C_2,
\end{aligned}$$

where C_1 and C_2 are defined in Appendix A.1. This completes the proof. ■

B.2.2 Time-varying Step-size

In this subsection, we consider the time-varying step-size case and begin with a property of η_t.

Lemma 8 *Suppose that Assumption 6 holds. Then,* $\lim_{t\to\infty}\eta_t = 0$ *and* $\lim_{t\to\infty}\frac{1}{t+1}\sum_{k=0}^{t}\eta_k = 0$.

Proof of Lemma 8: From Assumption 6, we know that π_t will converge to π_∞, and thus η_t will converge to 0. Next, we will prove that $\lim_{t\to\infty}\frac{1}{t+1}\sum_{k=0}^{t}\eta_k = 0$. For any positive constant $c > 0$, there exists a positive integer $T(c)$, depending on c, such that $\forall t \geq T(c)$, we have $\eta_t < c$. Thus,

$$\frac{1}{t}\sum_{k=0}^{t-1}\eta_k = \frac{1}{t}\sum_{k=0}^{T(c)}\eta_k + \frac{1}{t}\sum_{k=T(c)+1}^{t-1}\eta_k \leq \frac{1}{t}\sum_{k=0}^{T(c)}\eta_k + \frac{t-1-T(c)}{t}c.$$

Let $t \to \infty$ on both sides of the above inequality. Then, we have

$$\lim_{t\to\infty}\frac{1}{t}\sum_{k=0}^{t-1}\eta_k \leq \lim_{t\to\infty}\frac{1}{t}\sum_{k=0}^{T(c)}\eta_k + \lim_{t\to\infty}\frac{t-1-T(c)}{t}c = c.$$

Since the above argument holds for arbitrary positive c, then $\lim_{t\to\infty}\frac{1}{t+1}\sum_{k=0}^{t}\eta_k = 0$. ∎

Recall the updates corresponding to the time-varying step-size case given in (3) and (4),

$$\Theta_{t+1} = W_t\Theta_t + \alpha_t W_t\Theta_t A(X_t)^\top + \alpha_t B(X_t),$$
$$\langle\theta\rangle_{t+1} = \langle\theta\rangle_t + \alpha_t A(X_t)\langle\theta\rangle_t + \alpha_t B(X_t)^\top\pi_{t+1}.$$

From (25), we get the update for Y_t with the time-varying step-size as follows:

$$\begin{aligned}Y_{t+L} &= W_{t:t+L-1}Y_t(I+\alpha_t A^\top(X_t))\cdots(I+\alpha_{t+L-1}A^\top(X_{t+L-1}))\\ &\quad + \alpha_{t+L-1}(I-\mathbf{1}_N\pi_{t+L}^\top)B(X_{t+L-1})\\ &\quad + \sum_{k=t}^{t+L-2}\alpha_k W_{k+1:t+L-1}(I-\mathbf{1}_N\pi_{k+1}^\top)B(X_k)\left(\Pi_{j=k+1}^{t+L-1}(I+\alpha_j A^\top(X_j))\right),\end{aligned}$$

and

$$Y_{t+L}^i = \left(\Pi_{k=t}^{t+L-1}(I+\alpha_k A(X_k))\right)\sum_{j=1}^{N}w_{t:t+L-1}^{ij}Y_t^j + \tilde{b}_{t+L}^i,$$

where

$$\begin{aligned}\tilde{b}_{t+L}^i &= \alpha_{t+L-1}(b^i(X_{t+L-1}) - B(X_{t+L-1})^\top\pi_{t+L})\\ &\quad + \sum_{k=t}^{t+L-2}\alpha_k\left(\Pi_{j=k+1}^{t+L-1}(I+\alpha_j A(X_j))\right)\sum_{j=1}^{N}w_{k+1:t+L-1}^{ij}(b^j(X_k) - B(X_k)^\top\pi_{k+1}).\end{aligned}$$

To prove the theorem, we need the following lemmas.

Lemma 9 *Suppose that Assumptions 1 and 2 hold and* $\{\mathbb{G}_t\}$ *is uniformly strongly connected by sub-sequences of length* L*. Given* α_t *and* T_2 *defined in Theorem 3, for all* $t \geq T_2L$*,*

$$\begin{aligned}&\sum_{i=1}^{N}\pi_t^i\|\theta_t^i - \langle\theta\rangle_t\|_2^2\\ &\leq \epsilon^{q_t-T_2}\sum_{i=1}^{N}\pi_{T_2L+m_t}^i\|\theta_{T_2L+m_t}^i - \langle\theta\rangle_{T_2L+m_t}\|_2^2 + \frac{\zeta_6}{1-\epsilon}\left(\epsilon^{\frac{q_t-1}{2}}\alpha_{m_t} + \alpha_{\lceil\frac{q_t-1}{2}\rceil L+m_t}\right)\\ &\leq \epsilon^{q_t-T_2}\sum_{i=1}^{N}\pi_{T_2L+m_t}^i\|\theta_{T_2L+m_t}^i - \langle\theta\rangle_{T_2L+m_t}\|_2^2 + \frac{\zeta_6}{1-\epsilon}\left(\alpha_0\epsilon^{\frac{q_t-1}{2}} + \alpha_{\lceil\frac{q_t-1}{2}\rceil L}\right),\end{aligned}$$

where ϵ *and* ζ_6 *are defined in* (6) *and* (16)*, respectively.*

Proof of Lemma 9: Similar to the proof of Lemma 4, we have

$$\|Y_{t+L}\|_{M_{t+L}}^2 = \sum_{i=1}^N \pi_{t+L}^i \| \left(\Pi_{k=t}^{t+L-1}(I+\alpha_k A(X_k))\right) \sum_{j=1}^N w_{t:t+L-1}^{ij} Y_t^j \|_2^2 \tag{56}$$

$$+ \sum_{i=1}^N \pi_{t+L}^i \|\tilde{b}_{t+L}^i\|_2^2 \tag{57}$$

$$+ 2\sum_{i=1}^N \pi_{t+L}^i (\tilde{b}_{t+L}^i)^\top \left(\Pi_{k=t}^{t+L-1}(I+\alpha_k A(X_k))\right) \sum_{j=1}^N w_{t:t+L-1}^{ij} Y_t^j. \tag{58}$$

By using Lemma 3, the item given by (56) can be bounded as follows:

$$\begin{aligned}
&\sum_{i=1}^N \pi_{t+L}^i \| \left(\Pi_{k=t}^{t+L-1}(I+\alpha_k A(X_k))\right) \sum_{j=1}^N w_{t:t+L-1}^{ij} Y_t^j \|_2^2 \\
&\le \Pi_{k=t}^{t+L-1}(1+\alpha_k A_{\max})^2 \left[\sum_{i=1}^N \pi_t^i \|Y_t^i\|_2^2 - \frac{1}{2}\sum_{i=1}^N \pi_{t+L}^i \sum_{j=1}^N \sum_{l=1}^N w_{t:t+L-1}^{ij} w_{t:t+L-1}^{il} \|Y_t^j - Y_t^l\|_2^2\right] \\
&\le \Pi_{k=t}^{t+L-1}(1+\alpha_k A_{\max})^2 (1 - \frac{\pi_{\min}\beta^{2L}}{2\delta_{\max}}) \sum_{i=1}^N \pi_t^i \|Y_t^i\|_2^2.
\end{aligned} \tag{59}$$

Since $\|b^i(X_t) - B(X_t)^\top \pi_{t+1}\|_2 \le 2b_{\max}$ holds for all i, then

$$\begin{aligned}
&\|\tilde{b}_{t+L}^i\|_2 \\
&\le \alpha_{t+L-1} \|(b^i(X_{t+L-1}) - B(X_{t+L-1})^\top \pi_{t+L})\|_2 \\
&\quad + \sum_{k=t}^{t+L-2} \alpha_k \| \left(\Pi_{j=k+1}^{t+L-1}(I+\alpha_j A(X_j))\right) \|_2 \sum_{j=1}^N w_{k+1:t+L-1}^{ij} \|(b^j(X_k) - B(X_k)^\top \pi_{k+1})\|_2 \\
&\le 2b_{\max} \left[\alpha_{t+L-1} + \sum_{k=t}^{t+L-2} \alpha_k \left(\Pi_{j=k+1}^{t+L-1}(1+\alpha_j A_{\max})\right)\right].
\end{aligned}$$

Then, we can bound the item given by (57) as follows:

$$\sum_{i=1}^N \pi_{t+L}^i \|\tilde{b}_{t+L}^i\|_2^2 \le 4b_{\max}^2 \left(\alpha_{t+L-1} + \sum_{k=t}^{t+L-2} \alpha_k \left(\Pi_{j=k+1}^{t+L-1}(1+\alpha_j A_{\max})\right)\right)^2. \tag{60}$$

As for the item given by (58), we have

$$\begin{aligned}
&2\sum_{i=1}^N \pi_{t+L}^i (\tilde{b}_{t+L}^i)^\top \left(\Pi_{k=t}^{t+L-1}(I+\alpha_k A(X_k))\right) \sum_{j=1}^N w_{t:t+L-1}^{ij} Y_t^j \\
&\le 2\sum_{i=1}^N \pi_{t+L}^i \|\tilde{b}_{t+L}^i\|_2 \|\Pi_{k=t}^{t+L-1}(I+\alpha_k A(X_k))\|_2 \sum_{j=1}^N w_{t:t+L-1}^{ij} \|Y_t^j\|_2 \\
&\le 2b_{\max} \left(\alpha_{t+L-1} + \sum_{k=t}^{t+L-2} \alpha_k \left(\Pi_{j=k+1}^{t+L-1}(1+\alpha_j A_{\max})\right)\right) \left(\Pi_{k=t}^{t+L-1}(I+\alpha_k A_{\max})\right) \times \\
&\left(\sum_{i=1}^N \pi_t^i \|Y_t^i\|_2^2 + 1\right).
\end{aligned} \tag{61}$$

776 From (59)–(61), we have

$$
\begin{aligned}
&\|Y_{t+L}\|^2_{M_{t+L}} \\
&\le \Pi_{k=t}^{t+L-1}(1+\alpha_k A_{\max})^2(1-\frac{\pi_{\min}\beta^{2L}}{2\delta_{\max}})\sum_{i=1}^N \pi_t^i \|Y_t^i\|_2^2 \\
&\quad + 4b_{\max}^2\left(\alpha_{t+L-1} + \sum_{k=t}^{t+L-2}\alpha_k\left(\Pi_{j=k+1}^{t+L-1}(1+\alpha_j A_{\max})\right)\right)^2 \\
&\quad + 2b_{\max}\left(\alpha_{t+L-1} + \sum_{k=t}^{t+L-2}\alpha_k\left(\Pi_{j=k+1}^{t+L-1}(1+\alpha_j A_{\max})\right)\right)\left(\Pi_{k=t}^{t+L-1}(I+\alpha_k A_{\max})\right) \\
&\quad \times\left(\sum_{i=1}^N \pi_t^i\|Y_t^i\|_2^2 + 1\right) \\
&= \Bigg(2b_{\max}\left(\alpha_{t+L-1} + \sum_{k=t}^{t+L-2}\alpha_k\left(\Pi_{j=k+1}^{t+L-1}(1+\alpha_j A_{\max})\right)\right)\left(\Pi_{k=t}^{t+L-1}(I+\alpha_k A_{\max})\right) \\
&\quad + \Pi_{k=t}^{t+L-1}(1+\alpha_k A_{\max})^2(1-\frac{\pi_{\min}\beta^{2L}}{2\delta_{\max}})\Bigg)\|Y_t\|^2_{M_t} \\
&\quad + 4b_{\max}^2\left(\alpha_{t+L-1} + \sum_{k=t}^{t+L-2}\alpha_k\left(\Pi_{j=k+1}^{t+L-1}(1+\alpha_j A_{\max})\right)\right)^2 \\
&\quad + 2b_{\max}\left(\alpha_{t+L-1} + \sum_{k=t}^{t+L-2}\alpha_k\left(\Pi_{j=k+1}^{t+L-1}(1+\alpha_j A_{\max})\right)\right)\left(\Pi_{k=t}^{t+L-1}(I+\alpha_k A_{\max})\right) \\
&= \epsilon_t\|Y_t\|^2_{M_t} + 4b_{\max}^2\left(\alpha_{t+L-1} + \sum_{k=t}^{t+L-2}\alpha_k\left(\Pi_{j=k+1}^{t+L-1}(1+\alpha_j A_{\max})\right)\right)^2 \\
&\quad + 2b_{\max}\left(\alpha_{t+L-1} + \sum_{k=t}^{t+L-2}\alpha_k\left(\Pi_{j=k+1}^{t+L-1}(1+\alpha_j A_{\max})\right)\right)\left(\Pi_{k=t}^{t+L-1}(I+\alpha_k A_{\max})\right),
\end{aligned}
$$

777 where

$$
\begin{aligned}
\epsilon_t &= 2b_{\max}\left(\alpha_{t+L-1} + \sum_{k=t}^{t+L-2}\alpha_k\left(\Pi_{j=k+1}^{t+L-1}(1+\alpha_j A_{\max})\right)\right)\left(\Pi_{k=t}^{t+L-1}(I+\alpha_k A_{\max})\right) \\
&\quad + \Pi_{k=t}^{t+L-1}(1+\alpha_k A_{\max})^2(1-\frac{\pi_{\min}\beta^{2L}}{2\delta_{\max}}).
\end{aligned}
$$

778 Since for all $t \ge T_2 L$, we have $\alpha_t \le \alpha$, then for $t \ge T_2 L$ we have $0 \le \epsilon_t \le \epsilon \le 1$ and

$$
\begin{aligned}
\alpha_{t+L-1} + \sum_{k=t}^{t+L-2}\alpha_k\left(\Pi_{j=k+1}^{t+L-1}(1+\alpha_j A_{\max})\right) &\le \sum_{k=t}^{t+L-1}\alpha_k(1+\alpha A_{\max})^{t+L-k-1} \\
&\le (1+\alpha A_{\max})^{L-1}\sum_{k=t}^{t+L-1}\alpha_k.
\end{aligned}
$$

779 Since we have $\sum_{k=t}^{t+L-1}\alpha_k \le L\alpha_t \le L\alpha$. Then, we can get

$$
\begin{aligned}
&\|Y_{t+L}\|^2_{M_{t+L}} \\
&\le \epsilon\|Y_t\|^2_{M_t} + 4b_{\max}^2(1+\alpha A_{\max})^{2L-2}\left(\sum_{k=t}^{t+L-1}\alpha_k\right)^2 + 2b_{\max}(1+\alpha A_{\max})^{2L-1}\left(\sum_{k=t}^{t+L-1}\alpha_k\right) \\
&\le \epsilon\|Y_t\|^2_{M_t} + \left(4b_{\max}^2\alpha L^2(1+\alpha A_{\max})^{2L-2} + 2b_{\max}L(1+\alpha A_{\max})^{2L-1}\right)\alpha_t \\
&\le \epsilon\|Y_t\|^2_{M_t} + \zeta_6\alpha_t,
\end{aligned}
$$

where ϵ and ζ_6 are defined in (6) and (16) respectively. Then,

$$\begin{aligned}
&\|Y_{t+L}\|_{M_{t+L}}^2\\
&\leq \epsilon\|Y_t\|_{M_t}^2 + \zeta_6\alpha_t\\
&\leq \epsilon^{q_{t+L}-T_2}\|Y_{m_t+T_2L}\|_{M_{m_t+T_2L}}^2 + \zeta_6\sum_{k=T_2}^{q_t}\epsilon^{q_t-k}\alpha_{kL+m_t}\\
&\leq \epsilon^{q_{t+L}-T_2}\|Y_{T_2L+m_t}\|_{M_{T_2L+m_t}}^2 + \zeta_6\left(\sum_{k=0}^{\lfloor\frac{q_t}{2}\rfloor}\epsilon^{q_t-k}\alpha_{kL+m_t} + \sum_{k=\lceil\frac{q_t}{2}\rceil}^{q_t}\epsilon^{q_t-k}\alpha_{kL+m_t}\right)\\
&\leq \epsilon^{q_{t+L}-T_2}\|Y_{T_2L+m_t}\|_{M_{T_2L+m_t}}^2 + \frac{\zeta_6}{1-\epsilon}\left(\epsilon^{\frac{q_t}{2}}\alpha_{m_t} + \alpha_{\lceil\frac{q_t}{2}\rceil L+m_t}\right),
\end{aligned}$$

which implies

$$\begin{aligned}
&\sum_{i=1}^N \pi_t^i\|\theta_t^i - \langle\theta\rangle_t\|_2^2\\
&\leq \epsilon^{q_t-T_2}\sum_{i=1}^N \pi_{T_2L+m_t}^i\|\theta_{T_2L+m_t}^i - \langle\theta\rangle_{T_2L+m_t}\|_2^2 + \frac{\zeta_6}{1-\epsilon}\left(\epsilon^{\frac{q_t-1}{2}}\alpha_{m_t} + \alpha_{\lceil\frac{q_t-1}{2}\rceil L+m_t}\right)\\
&\leq \epsilon^{q_t-T_2}\sum_{i=1}^N \pi_{T_2L+m_t}^i\|\theta_{T_2L+m_t}^i - \langle\theta\rangle_{T_2L+m_t}\|_2^2 + \frac{\zeta_6}{1-\epsilon}\left(\alpha_0\epsilon^{\frac{q_t-1}{2}} + \alpha_{\lceil\frac{q_t-1}{2}\rceil L}\right).
\end{aligned}$$

This completes the proof. ∎

Lemma 10 *Suppose that Assumptions 2 and 3 hold. When the step-size α_t and corresponding mixing time $\tau(\alpha_t)$ satisfy*

$$0 < \alpha_t\tau(\alpha_t) < \frac{\log 2}{A_{\max}},$$

we have for any $t \geq T_2L$,

$$\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 \leq 2A_{\max}\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k + 2b_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k, \tag{62}$$

$$\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 \leq 6A_{\max}\|\langle\theta\rangle_t\|_2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k + 5b_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k, \tag{63}$$

$$\begin{aligned}
\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2 &\leq 72\alpha_{t-\tau(\alpha_t)}^2\tau^2(\alpha_t)A_{\max}^2\|\langle\theta\rangle_t\|_2^2 + 50\alpha_{t-\tau(\alpha_t)}^2\tau^2(\alpha_t)b_{\max}^2\\
&\leq 8\|\langle\theta\rangle_t\|_2^2 + \frac{6b_{\max}^2}{A_{\max}^2}.
\end{aligned} \tag{64}$$

Proof of Lemma 10: Recall the update of $\langle\theta\rangle_t$ in (4):

$$\langle\theta\rangle_{t+1} = \langle\theta\rangle_t + \alpha_t A(X_t)\langle\theta\rangle_t + \alpha_t B(X_t)^\top\pi_{t+1}.$$

Then, we have

$$\|\langle\theta\rangle_{t+1}\|_2 \leq \|\langle\theta\rangle_t\|_2 + \alpha_t A_{\max}\|\langle\theta\rangle_t\|_2 + \alpha_t b_{\max} \leq (1+\alpha_t A_{\max})\|\langle\theta\rangle_t\|_2 + \alpha_t b_{\max}.$$

788 Similar to the proof of Lemma 5, for all $u \in [t-\tau(\alpha_t), t]$, we have

$$
\begin{aligned}
&\|\langle\theta\rangle_u\|_2 \\
&\le \Pi_{k=t-\tau(\alpha_t)}^{u-1}(1+\alpha_k A_{\max})\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 + b_{\max}\sum_{k=t-\tau(\alpha_t)}^{u-1}\alpha_k \Pi_{l=k+1}^{u-1}(1+\alpha_l A_{\max}) \\
&\le \exp\{\sum_{k=t-\tau(\alpha_t)}^{u-1}\alpha_k A_{\max}\}\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 + b_{\max}\sum_{k=t-\tau(\alpha_t)}^{u-1}\alpha_k \exp\{\sum_{l=k+1}^{u-1}\alpha_l A_{\max}\} \\
&\le \exp\{\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)A_{\max}\}\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 + b_{\max}\sum_{k=t-\tau(\alpha_t)}^{u-1}\alpha_k \exp\{\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)A_{\max}\} \\
&\le 2\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 + 2b_{\max}\sum_{k=t-\tau(\alpha_t)}^{u-1}\alpha_k,
\end{aligned}
$$

789 where we use $\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)A_{\max} \le \log 2 < \frac{1}{3}$ in the last inequality. Thus, for all $t \ge T_2 L$, we can
790 get (62) as follows:

$$
\begin{aligned}
&\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 \\
&\le \sum_{k=t-\tau(\alpha_t)}^{t-1}\|\langle\theta\rangle_{k+1} - \langle\theta\rangle_k\|_2 \\
&\le A_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k\|\langle\theta\rangle_k\|_2 + b_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\le A_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k\left(2\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 + 2b_{\max}\sum_{l=t-\tau(\alpha_t)}^{k-1}\alpha_l\right) + b_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\le 2A_{\max}\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k + \left(2A_{\max}\tau(\alpha_t)\alpha_{t-\tau(\alpha_t)} + 1\right)b_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\le 2A_{\max}\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k + \frac{5}{3}b_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\le 2A_{\max}\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k + 2b_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k.
\end{aligned}
$$

791 Moreover, by using the above inequality, we can get (63) for all $t \ge T_2 L$ as follows:

$$
\begin{aligned}
&\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 \\
&\le 2A_{\max}\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k + \frac{5}{3}b_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\le 2A_{\max}\tau(\alpha_t)\alpha_{t-\tau(\alpha_t)}\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 + 2A_{\max}\|\langle\theta\rangle_t\|_2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\quad + \frac{5}{3}b_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\le 6A_{\max}\|\langle\theta\rangle_t\|_2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k + 5b_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k.
\end{aligned}
$$

Next, by using (63) and the inequality $(x+y)^2 \leq 2x^2 + y^2$ for all x, y, we can get (64) as follows:

$$\begin{aligned}\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2 &\leq 72A_{\max}^2\|\langle\theta\rangle_t\|_2^2(\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k)^2 + 50b_{\max}^2(\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k)^2 \\ &\leq 72\alpha_{t-\tau(\alpha_t)}^2\tau^2(\alpha_t)A_{\max}^2\|\langle\theta\rangle_t\|_2^2 + 50\alpha_{t-\tau(\alpha_t)}^2\tau^2(\alpha_t)b_{\max}^2 \\ &\leq 8\|\langle\theta\rangle_t\|_2^2 + 6(\frac{b_{\max}}{A_{\max}})^2,\end{aligned}$$

where we use $\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)A_{\max} < \frac{1}{3}$ in the last inequality. ■

Lemma 11 *Suppose that Assumptions 2–6 hold and $\{\mathbb{G}_t\}$ is uniformly strongly connected. When*

$$0 < \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t) < \frac{\log 2}{A_{\max}},$$

we have for any $t \geq T_2L$,

$$\begin{aligned}&\left|\mathbf{E}\left[(\langle\theta\rangle_t - \theta^*)^\top(P+P^\top)\left(A(X_t)\langle\theta\rangle_t + B(X_t)^\top\pi_{t+1} - A\langle\theta\rangle_t - b\right) \mid \mathcal{F}_{t-\tau(\alpha_t)}\right]\right| \\ &\leq \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\left(72 + 456A_{\max}^2 + 84A_{\max}b_{\max}\right)\mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\ &\quad + \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\Bigg[2 + 4\|\theta^*\|_2^2 + \frac{48b_{\max}^2}{A_{\max}^2} + 152\left(b_{\max} + A_{\max}\|\theta^*\|_2\right)^2 + 12A_{\max}b_{\max} \\ &\quad + 48A_{\max}b_{\max}\left(\frac{b_{\max}}{A_{\max}} + 1\right)^2 + 87b_{\max}^2\Bigg] \\ &\quad + 2\gamma_{\max}\eta_{t+1}\sqrt{N}b_{\max}\left(1 + 9\mathbf{E}\left[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha)}\right] + \frac{6b_{\max}^2}{A_{\max}^2} + \|\theta^*\|_2^2\right).\end{aligned}$$

Proof of Lemma 11: Note that for all $t \geq T_2L$, we have

$$\begin{aligned}&|\mathbf{E}[(\langle\theta\rangle_t - \theta^*)^\top(P+P^\top)(A(X_t)\langle\theta\rangle_t + B(X_t)^\top\pi_{t+1} - A\langle\theta\rangle_t - b) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\ &\leq |\mathbf{E}[(\langle\theta\rangle_t - \theta^*)^\top(P+P^\top)(A(X_t)-A)\langle\theta\rangle_t \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\ &\quad + |\mathbf{E}[(\langle\theta\rangle_t - \theta^*)^\top(P+P^\top)(B(X_t)^\top\pi_{t+1} - b) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\ &\leq |\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top(P+P^\top)(A(X_t)-A)\langle\theta\rangle_{t-\tau(\alpha_t)} \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \qquad (65) \\ &\quad + |\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top(P+P^\top)(A(X_t)-A)(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \qquad (66) \\ &\quad + |\mathbf{E}[(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)})^\top(P+P^\top)(A(X_t)-A)\langle\theta\rangle_{t-\tau(\alpha_t)} \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \qquad (67) \\ &\quad + |\mathbf{E}[(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)})^\top(P+P^\top)(A(X_t)-A)(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \qquad (68) \\ &\quad + |\mathbf{E}[(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)})^\top(P+P^\top)(B(X_t)^\top\pi_{t+1} - b) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \qquad (69) \\ &\quad + |\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top(P+P^\top)(B(X_t)^\top\pi_{t+1} - b) \mid \mathcal{F}_{t-\tau(\alpha_t)}]|. \qquad (70)\end{aligned}$$

Similar to the proof of Lemma 6, by using the mixing time in Assumption 3, we can get the bound for (65) and (70) for all $t \geq T_2L$:

$$\begin{aligned}&|\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top(P+P^\top)(A(X_t)-A)\langle\theta\rangle_{t-\tau(\alpha_t)} \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\ &\leq |(\langle\theta\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top(P+P^\top)\mathbf{E}[A(X_t) - A \mid \mathcal{F}_{t-\tau(\alpha_t)}]\langle\theta\rangle_{t-\tau(\alpha_t)}| \\ &\leq 2\alpha_t\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha_t)} - \theta^*\|_2\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\ &\leq \alpha_t\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha_t)} - \theta^*\|_2^2 + \|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\ &\leq \alpha_t\gamma_{\max}\mathbf{E}[2\|\theta^*\|_2^2 + 3\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\ &\leq 6\alpha_t\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] + 6\alpha_t\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] + 2\alpha_t\gamma_{\max}\|\theta^*\|_2^2 \\ &\leq 54\alpha_t\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] + 36\alpha_t\gamma_{\max}(\frac{b_{\max}}{A_{\max}})^2 + 2\alpha_t\gamma_{\max}\|\theta^*\|_2^2, \qquad (71)\end{aligned}$$

where in the last inequality, we use (64) from Lemma 10.

$$
\begin{aligned}
&|\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha_t)}-\theta^*)^\top(P+P^\top)(B(X_t)^\top\pi_{t+1}-b)\mid\mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le |\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha_t)}-\theta^*)^\top(P+P^\top)(\sum_{i=1}^N\pi_{t+1}^i(b^i(X_t)-b^i)+\sum_{i=1}^N(\pi_{t+1}^i-\pi_\infty^i)b^i)\mid\mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le |(\langle\theta\rangle_{t-\tau(\alpha_t)}-\theta^*)^\top(P+P^\top)(\sum_{i=1}^N\pi_{t+1}^i\mathbf{E}[b^i(X_t)-b^i\mid\mathcal{F}_{t-\tau(\alpha_t)}]+\sum_{i=1}^N(\pi_{t+1}^i-\pi_\infty^i)b^i)| \\
&\le 2\gamma_{\max}(\alpha_t+\eta_{t+1}\sqrt{N}b_{\max})\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha_t)}-\theta^*\|_2\mid\mathcal{F}_{t-\tau(\alpha_t)}] \\
&\le 2\gamma_{\max}(\alpha_t+\eta_{t+1}\sqrt{N}b_{\max})\left(1+\frac{1}{2}\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]+\frac{1}{2}\|\theta^*\|_2^2\right) \\
&\le 2\gamma_{\max}(\alpha_t+\eta_{t+1}\sqrt{N}b_{\max})\left(1+\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha_t)}-\langle\theta\rangle_t\|_2^2+\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]+\|\theta^*\|_2^2\right) \\
&\le 2\gamma_{\max}(\alpha_t+\eta_{t+1}\sqrt{N}b_{\max})\left(1+9\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]+6(\frac{b_{\max}}{A_{\max}})^2+\|\theta^*\|_2^2\right), \qquad (72)
\end{aligned}
$$

where in the last inequality we use (64).

Next, by using Assumption 2, (62) and (64), we have

$$
\begin{aligned}
&|\mathbf{E}[(\langle\theta\rangle_{t-\tau(\alpha_t)}-\theta^*)^\top(P+P^\top)(A(X_t)-A)(\langle\theta\rangle_t-\langle\theta\rangle_{t-\tau(\alpha_t)})\mid\mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le 4\gamma_{\max}A_{\max}\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha_t)}-\theta^*\|_2\|\langle\theta\rangle_t-\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2\mid\mathcal{F}_{t-\tau(\alpha_t)}] \\
&\le 4\gamma_{\max}A_{\max}\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2\|\langle\theta\rangle_t-\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2+\|\theta^*\|_2\|\langle\theta\rangle_t-\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2\mid\mathcal{F}_{t-\tau(\alpha_t)}] \\
&\le 8\gamma_{\max}A_{\max}^2\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k+8\gamma_{\max}A_{\max}b_{\max}\|\theta^*\|_2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\quad+8\gamma_{\max}A_{\max}^2\left(\frac{b_{\max}}{A_{\max}}+\|\theta^*\|_2\right)\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2\mid\mathcal{F}_{t-\tau(\alpha_t)}]\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\le \gamma_{\max}A_{\max}^2\left(12\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]+8\left(\frac{b_{\max}}{A_{\max}}+\|\theta^*\|_2\right)^2\right)\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\le 24\gamma_{\max}A_{\max}^2\mathbf{E}[\|\langle\theta\rangle_t-\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\quad+24\gamma_{\max}A_{\max}^2\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\quad+8\gamma_{\max}A_{\max}^2\left(\frac{b_{\max}}{A_{\max}}+\|\theta^*\|_2\right)^2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\le \gamma_{\max}\left(216A_{\max}^2\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]+152\left(b_{\max}+A_{\max}\|\theta^*\|_2\right)^2\right)\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k. \qquad (73)
\end{aligned}
$$

In additional, as for the bound of (67), by using (62) and (64), we have

$$
\begin{aligned}
&|\mathbf{E}[(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)})^\top (P+P^\top)(A(X_t)-A)\langle\theta\rangle_{t-\tau(\alpha_t)} \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le 4\gamma_{\max} A_{\max} \mathbf{E}[\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 \|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\
&\le 8\gamma_{\max} A_{\max} \mathbf{E}[A_{\max}\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2 + b_{\max}\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 4\gamma_{\max} A_{\max} \left((2A_{\max}+b_{\max})\mathbf{E}[\|\langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] + b_{\max}\right) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 8\gamma_{\max} A_{\max}(2A_{\max}+b_{\max})\mathbf{E}[\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 8\gamma_{\max} A_{\max}(2A_{\max}+b_{\max})\mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 4\gamma_{\max} A_{\max} b_{\max} \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 72\, \gamma_{\max} A_{\max}(2A_{\max}+b_{\max})\mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 48\gamma_{\max} A_{\max} b_{\max} (\frac{b_{\max}}{A_{\max}} + 1)^2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k. \qquad (74)
\end{aligned}
$$

Moreover, by using (64), we can get the bound for (68) as follows:

$$
\begin{aligned}
&|\mathbf{E}[(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)})^\top (P+P^\top)(A(X_t)-A)(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le 4\gamma_{\max} A_{\max} \mathbf{E}[\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le 4\gamma_{\max} A_{\max} \mathbf{E}[72A_{\max}^2 \|\langle\theta\rangle_t\|_2^2 + 50b_{\max}^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}](\sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k)^2 \\
&\le 96A_{\max}^2 \gamma_{\max} \mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k + 67b_{\max}^2 \gamma_{\max} \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k. \qquad (75)
\end{aligned}
$$

Finally, we can get the bound of (69) by using (63):

$$
\begin{aligned}
&|\mathbf{E}[(\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)})(P+P^\top)(B(X_t)^\top \pi_{t+1} - b) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le 4\gamma_{\max} b_{\max} \mathbf{E}[\|\langle\theta\rangle_t - \langle\theta\rangle_{t-\tau(\alpha_t)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\
&\le 4\gamma_{\max} b_{\max} \mathbf{E}[6A_{\max}\|\langle\theta\rangle_t\|_2 + 5b_{\max} \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le \gamma_{\max} \left(12A_{\max} b_{\max} \mathbf{E}[\|\langle\theta\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] + 12A_{\max} b_{\max} + 20b_{\max}^2\right) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k. \qquad (76)
\end{aligned}
$$

Then, by using (71)–(76), we have

$$
\begin{aligned}
&|\mathbf{E}[(\langle\theta\rangle_t-\theta^*)^\top(P+P^\top)(A(X_t)\langle\theta\rangle_t+B(X_t)^\top\pi_{t+1}-A\langle\theta\rangle_t-b)\mid\mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le 54\alpha_t\gamma_{\max}\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]+36\alpha_t\gamma_{\max}(\frac{b_{\max}}{A_{\max}})^2+2\alpha_t\gamma_{\max}\|\theta^*\|_2^2 \\
&\quad+216\gamma_{\max}A_{\max}^2\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\quad+152\gamma_{\max}\left(b_{\max}+A_{\max}\|\theta^*\|_2\right)^2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k+67b_{\max}^2\gamma_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\quad+12\,\gamma_{\max}A_{\max}(20A_{\max}+7b_{\max})\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\quad+48\gamma_{\max}A_{\max}b_{\max}(\frac{b_{\max}}{A_{\max}}+1)^2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k+(12A_{\max}b_{\max}+20b_{\max}^2)\gamma_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k \\
&\quad+2\gamma_{\max}(\alpha_t+\eta_{t+1}\sqrt{N}b_{\max})\left(1+9\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]+6(\frac{b_{\max}}{A_{\max}})^2+\|\theta^*\|_2^2\right) \\
&\le \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\left(72+456A_{\max}^2+84A_{\max}b_{\max}\right)\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}] \\
&\quad+\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\Big[2+4\|\theta^*\|_2^2+48(\frac{b_{\max}}{A_{\max}})^2+152\left(b_{\max}+A_{\max}\|\theta^*\|_2\right)^2+12A_{\max}b_{\max} \\
&\quad+48A_{\max}b_{\max}(\frac{b_{\max}}{A_{\max}}+1)^2+87b_{\max}^2\Big] \\
&\quad+2\gamma_{\max}\eta_{t+1}\sqrt{N}b_{\max}\left(1+9\mathbf{E}[\|\langle\theta\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha)}]+6(\frac{b_{\max}}{A_{\max}})^2+\|\theta^*\|_2^2\right),
\end{aligned}
$$

where we use $\alpha_t\le\alpha_{t-\tau(\alpha_t)}$ from Assumption 5 and $\tau(\alpha_t)\ge 1$ in the last inequality. This completes the proof. ■

Lemma 12 *Under Assumptions 1–6, when the* $\tau(\alpha_t)\alpha_{t-\tau(\alpha_t)}\le\min\{\frac{\log 2}{A_{\max}},\ \frac{0.1}{\zeta_5\gamma_{\max}}\}$*, we have for any* $t\ge T_2L$*,*

$$
\begin{aligned}
\mathbf{E}\left[\|\langle\theta\rangle_t-\theta^*\|_2^2\right]\le\frac{T_2L}{t}\frac{\gamma_{\max}}{\gamma_{\min}}\mathbf{E}[\|\langle\theta\rangle_{T_2L}-\theta^*\|_2^2]+\frac{\zeta_7\alpha_0C\log^2(\frac{t}{\alpha_0})}{t}\frac{\gamma_{\max}}{\gamma_{\min}} \\
+\alpha_0\zeta_4\frac{\gamma_{\max}}{\gamma_{\min}}\frac{\sum_{l=T_2L}^t\eta_l}{t},
\end{aligned}
$$

where T_2 *is defined in Appendix A.1, and* ζ_4*,* ζ_5*,* ζ_7 *are defined in* (14)*,* (15)*,* (17)*, respectively.*

Proof of Lemma 12: Recall the update of $\langle\theta\rangle_t$ in (4):

$$
\langle\theta\rangle_{t+1}=\langle\theta\rangle_t+\alpha_tA(X_t)\langle\theta\rangle_t+\alpha_tB(X_t)^\top\pi_{t+1}.
$$

Note that $\mathbf{E}[\|\langle\theta\rangle_t\|_2^2] \le 2\mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] + 2\|\theta^*\|_2^2 \le \frac{2}{\gamma_{\min}}\mathbf{E}[H(\langle\theta\rangle_t)] + 2\|\theta^*\|_2^2$, then from (52) and Lemma 11, for $t \ge T_2 L$ we have

$$
\begin{aligned}
&\mathbf{E}[H(\langle\theta\rangle_{t+1})] \\
&\le \mathbf{E}[H(\langle\theta\rangle_t)] - \alpha_t \mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] + \alpha_t^2 A_{\max}^2 \gamma_{\max} \mathbf{E}[\|\langle\theta\rangle_t\|_2^2] + \alpha_t^2 b_{\max}^2 \gamma_{\max} \\
&\quad + 2\alpha_t^2 A_{\max} b_{\max} \gamma_{\max} \mathbf{E}[\|\langle\theta\rangle_t\|_2] \\
&\quad + \alpha_t \mathbf{E}[(\langle\theta\rangle_t - \theta^*)^\top (P + P^\top)(A(X_t)\langle\theta\rangle_t + B(X_t)^\top \pi_{t+1} - A\langle\theta\rangle_t - b)] \\
&\le \mathbf{E}[H(\langle\theta\rangle_t)] - \alpha_t \mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] + 2\alpha_t^2 A_{\max}^2 \gamma_{\max} \mathbf{E}[\|\langle\theta\rangle_t\|_2^2] + 2\alpha_t^2 b_{\max}^2 \gamma_{\max} \\
&\quad + \alpha_t \alpha_{t-\tau(\alpha_t)} \tau(\alpha_t) \gamma_{\max} \left(72 + 456 A_{\max}^2 + 84 A_{\max} b_{\max}\right) \mathbf{E}[\|\langle\theta\rangle_t\|_2^2] \\
&\quad + \alpha_t \alpha_{t-\tau(\alpha_t)} \tau(\alpha_t) \gamma_{\max} \Bigg[2 + 4\|\theta^*\|_2^2 + 48(\frac{b_{\max}}{A_{\max}})^2 + 152\left(b_{\max} + A_{\max}\|\theta^*\|_2\right)^2 \\
&\quad + 12 A_{\max} b_{\max} + 48 A_{\max} b_{\max} (\frac{b_{\max}}{A_{\max}} + 1)^2 + 87 b_{\max}^2\Bigg] \\
&\quad + 2\alpha_t \gamma_{\max} \eta_{t+1} \sqrt{N} b_{\max} \left(1 + 9\mathbf{E}[\|\langle\theta\rangle_t\|_2^2] + 6(\frac{b_{\max}}{A_{\max}})^2 + \|\theta^*\|_2^2\right) \\
&\le \mathbf{E}[H(\langle\theta\rangle_t)] + 2\alpha_t \alpha_{t-\tau(\alpha_t)} \tau(\alpha_t) \gamma_{\max} \left(72 + 458 A_{\max}^2 + 84 A_{\max} b_{\max}\right) \|\theta^*\|_2^2 \\
&\quad + \left(-\alpha_t + 2\alpha_t \alpha_{t-\tau(\alpha_t)} \tau(\alpha_t) \gamma_{\max} \left(72 + 458 A_{\max}^2 + 84 A_{\max} b_{\max}\right)\right) \mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] \\
&\quad + \alpha_t \alpha_{t-\tau(\alpha_t)} \tau(\alpha_t) \gamma_{\max} \Bigg[2 + 4\|\theta^*\|_2^2 + 48(\frac{b_{\max}}{A_{\max}})^2 + 152\left(b_{\max} + A_{\max}\|\theta^*\|_2\right)^2 \\
&\quad + 12 A_{\max} b_{\max} + 48 A_{\max} b_{\max} (\frac{b_{\max}}{A_{\max}} + 1)^2 + 89 b_{\max}^2\Bigg] \\
&\quad + 2\alpha_t \gamma_{\max} \eta_{t+1} \sqrt{N} b_{\max} \left(1 + 18\mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] + 6(\frac{b_{\max}}{A_{\max}})^2 + 19\|\theta^*\|_2^2\right) \\
&\le \mathbf{E}[H(\langle\theta\rangle_t)] + \left(-\alpha_t + \alpha_t \alpha_{t-\tau(\alpha_t)} \tau(\alpha_t) \gamma_{\max} \zeta_5 + 36\alpha_t \gamma_{\max} \eta_{t+1} \sqrt{N} b_{\max}\right) \mathbf{E}[\|\langle\theta\rangle_t - \theta^*\|_2^2] \\
&\quad + \alpha_t \alpha_{t-\tau(\alpha_t)} \tau(\alpha_t) \gamma_{\max} \zeta_7 + \alpha_t \gamma_{\max} \eta_{t+1} \zeta_4,
\end{aligned}
$$

where ζ_4, ζ_5 and ζ_7 are defined in (14), (15) and (17), respectively. Moreover, from $\alpha_t = \frac{\alpha_0}{t+1}$, $\alpha_0 \ge \frac{\gamma_{max}}{0.9}$ and the definition of T_2, we have for all $t \ge T_2 L$

$$
\begin{aligned}
\mathbf{E}[H(\langle\theta\rangle_{t+1})] &\le \left(1 - \frac{0.9\alpha_t}{\gamma_{\max}}\right) \mathbf{E}[H(\langle\theta\rangle_t)] + \alpha_t \gamma_{\max} \eta_{t+1} \zeta_4 + \alpha_t \alpha_{t-\tau(\alpha_t)} \tau(\alpha_t) \gamma_{\max} \zeta_7 \\
&\le \frac{t}{t+1} \mathbf{E}[H(\langle\theta\rangle_t)] + \alpha_0 \gamma_{\max} \zeta_4 \frac{\eta_{t+1}}{t+1} + \frac{\alpha_0^2 C \log(\frac{t+1}{\alpha_0}) \gamma_{\max} \zeta_7}{(t+1)(t - \tau(\alpha_t) + 1)} \\
&\le \frac{T_2 L}{t+1} \mathbf{E}[H(\langle\theta\rangle_{T_2 L})] + \alpha_0 \gamma_{\max} \zeta_4 \sum_{l=T_2 L}^{t} \frac{\eta_{l+1}}{l+1} \Pi_{u=l+1}^{t} \frac{u}{u+1} \\
&\quad + \alpha_0^2 \gamma_{\max} \zeta_7 \sum_{l=T_2 L}^{t} \frac{C \log(\frac{l+1}{\alpha_0})}{(l+1)(l - \tau(\alpha_l) + 1)} \Pi_{u=l+1}^{t} \frac{u}{u+1} \\
&\le \frac{T_2 L}{t+1} \mathbf{E}[H(\langle\theta\rangle_{T_2 L})] + \alpha_0 \gamma_{\max} \zeta_4 \frac{\sum_{l=T_2 L}^{t} \eta_{l+1}}{t+1} + \frac{\zeta_7 \alpha_0 \gamma_{\max} C \log^2(\frac{t+1}{\alpha_0})}{t+1} \\
&\le \frac{T_2 L}{t+1} \mathbf{E}[H(\langle\theta\rangle_{T_2 L})] + \alpha_0 \gamma_{\max} \zeta_4 \frac{\sum_{l=T_2 L}^{t+1} \eta_l}{t+1} + \frac{\zeta_7 \alpha_0 \gamma_{\max} C \log^2(\frac{t+1}{\alpha_0})}{t+1},
\end{aligned}
$$

where we use

$$
\sum_{l=T_2}^{t} \frac{2\alpha_0 \log(\frac{l+1}{\alpha_0})}{l+1} \le \log^2(\frac{t+1}{\alpha_0})
$$

to get the last inequality. Then, we can get the bound of $\mathbf{E}[\|\langle\theta\rangle_{t+1}-\theta^*\|_2^2]$ as follows

$$\begin{aligned}\mathbf{E}[\|\langle\theta\rangle_{t+1}-\theta^*\|_2^2] &\le \frac{1}{\gamma_{\min}}\mathbf{E}[H(\langle\theta\rangle_{t+1})] \\ &\le \frac{T_2L}{t+1}\frac{\gamma_{\max}}{\gamma_{\min}}\mathbf{E}[\|\langle\theta\rangle_{T_2L}-\theta^*\|_2^2] + \frac{\zeta_7\alpha_0 C\log^2(\frac{t+1}{\alpha_0})}{t+1}\frac{\gamma_{\max}}{\gamma_{\min}} \\ &\quad + \alpha_0\zeta_4\frac{\gamma_{\max}}{\gamma_{\min}}\frac{\sum_{l=T_2L}^{t+1}\eta_l}{t+1}.\end{aligned}$$

This completes the proof. ■

We are now in a position to prove the time-varying step-size case in Theorem 3.

Proof of Case 2) in Theorem 3: From Lemmas 9 and 12, for any $t\ge T_2L$, we have

$$\begin{aligned}\sum_{i=1}^N \pi_t^i\mathbf{E}[\|\theta_t^i-\theta^*\|_2^2] &\le 2\sum_{i=1}^N \pi_t^i\mathbf{E}[\|\theta_t^i-\langle\theta\rangle_t\|_2^2] + 2\mathbf{E}[\|\langle\theta\rangle_t-\theta^*\|_2^2] \\ &\le 2\epsilon^{q_t-T_2}\sum_{i=1}^N \pi_{T_2L+m_t}^i\mathbf{E}[\|\theta_{T_2L+m_t}^i-\langle\theta\rangle_{T_2L+m_t}\|_2^2] \\ &\quad + \frac{2T_2L}{t}\frac{\gamma_{\max}}{\gamma_{\min}}\mathbf{E}[\|\langle\theta\rangle_{T_2L}-\theta^*\|_2^2] + \frac{2\zeta_7\alpha_0 C\log^2(\frac{t}{\alpha_0})}{t}\frac{\gamma_{\max}}{\gamma_{\min}} \\ &\quad + 2\alpha_0\zeta_4\frac{\gamma_{\max}}{\gamma_{\min}}\frac{\sum_{l=T_2L}^{t}\eta_l}{t} + \frac{2\zeta_6}{1-\epsilon}(\alpha_0\epsilon^{\frac{q_t-1}{2}} + \alpha_{\lceil\frac{q_t-1}{2}\rceil L}) \\ &\le 2\epsilon^{q_t-T_2}\sum_{i=1}^N \pi_{LT_2+m_t}^i\mathbf{E}\left[\left\|\theta_{LT_2+m_t}^i-\langle\theta\rangle_{LT_2+m_t}\right\|_2^2\right] \\ &\quad + C_3\left(\alpha_0\epsilon^{\frac{q_t-1}{2}} + \alpha_{\lceil\frac{q_t-1}{2}\rceil L}\right) + \frac{1}{t}\left(C_4\log^2\left(\frac{t}{\alpha_0}\right) + C_5\sum_{k=LT_2}^t\eta_k + C_6\right),\end{aligned}$$

where C_3-C_6 are defined in Appendix A.1. This completes the proof. ■

B.3 Push-SA

In this subsection, we analyze the push-based distributed stochastic approximation algorithm (9) and provide the proofs of the results in Section 3. We begin with the proof of asymptotic performance.

Proof of Theorem 4: From Lemma 19, since $\bar\epsilon\in(0,1)$ and $\alpha_t=\frac{\alpha_0}{t}$, we have $\lim_{t\to\infty}\|\theta_{t+1}^i-\langle\tilde\theta\rangle_t\|_2=0$, which implies that all θ_{t+1}^i, $i\in\mathcal{V}$, will reach a consensus with $\langle\tilde\theta\rangle_t$. The update of $\langle\tilde\theta\rangle_t$ is given in (82), which can be treated as a single-agent linear stochastic approximation whose corresponding ODE is (10). In addition, from Theorem 5 and Lemma 20, $\lim_{\to\infty}\sum_{i=1}^N\mathbf{E}[\|\theta_{t+1}^i-\theta^*\|_2^2]=0$, it follows that θ_{t+1}^i will converge to θ^* in mean square for all $i\in\mathcal{V}$. ■

We now analyze the finite-time performance of (9).

Let $\hat W_t$ be the matrix whose ij-th entry is $\hat w_t^{ij}$. Then, from (9) we have

$$\begin{aligned}\theta_{t+1}^i = \frac{\tilde\theta_{t+1}^i}{y_{t+1}^i} &= \frac{\sum_{j=1}^N\hat w_t^{ij}(\tilde\theta_t^j+\alpha_tA(X_t)\theta_t^j+\alpha_tb^j(X_t))}{y_{t+1}^i} \\ &= \sum_{j=1}^N\frac{\hat w_t^{ij}y_t^j}{\sum_{k=1}^N\hat w_t^{ik}y_t^k}\left[\frac{\tilde\theta_t^j}{y_t^j}+\alpha_tA(X_t)\frac{\theta_t^j}{y_t^j}+\alpha_t\frac{b^j(X_t)}{y_t^j}\right] \\ &= \sum_{j=1}^N\tilde w_t^{ij}\left[\theta_t^j+\alpha_tA(X_t)\frac{\theta_t^j}{y_t^j}+\alpha_t\frac{b^j(X_t)}{y_t^j}\right],\end{aligned}\tag{77}$$

where $\tilde{w}_t^{ij} = \frac{\hat{w}_t^{ij} y_t^j}{\sum_{k=1}^N \hat{w}_t^{ik} y_t^k}$ and $\tilde{W}_t = [\tilde{w}_t^{ij}]$ is a row stochastic matrix, i.e.,

$$\sum_{j=1}^N \tilde{w}_t^{ij} = \frac{\sum_{j=1}^N \hat{w}_t^{ij} y_t^j}{\sum_{k=1}^N \hat{w}_t^{ik} y_t^k} = 1, \quad \forall i.$$

Let $\Theta_t = [\theta_t^1, \cdots, \theta_t^N]^\top$ and $\tilde{\Theta}_t = [\tilde{\theta}_t^1, \cdots, \tilde{\theta}_t^N]^\top$. Then (9) and (77) can be written as

$$\tilde{\Theta}_{t+1} = \hat{W}_t \left[\tilde{\Theta}_t + \alpha_t \begin{bmatrix} (\tilde{\theta}_t^1)^\top / y_t^1 \\ \cdots \\ (\tilde{\theta}_t^N)^\top / y_t^N \end{bmatrix} A(X_t)^\top + \alpha_t B(X_t) \right] \tag{78}$$

$$\Theta_{t+1} = \tilde{W}_t \left[\Theta_t + \alpha_t \begin{bmatrix} (\theta_t^1)^\top / y_t^1 \\ \cdots \\ (\theta_t^N)^\top / y_t^N \end{bmatrix} A(X_t)^\top + \alpha_t \begin{bmatrix} (b^1(X_t))^\top / y_t^1 \\ \cdots \\ (b^N(X_t))^\top / y_t^N \end{bmatrix} \right]. \tag{79}$$

Since each matrix $\tilde{W}_t = [\tilde{w}_t^{ij}]$ is stochastic, from Lemma 1, there exists a unique absolute probability sequence $\{\tilde{\pi}_t\}$ for the matrix sequence $\{\tilde{W}_t\}$ such that $\tilde{\pi}_t^i \geq \tilde{\pi}_{\min}$ for all $i \in \mathcal{V}$ and $t \geq 0$, with the constant $\tilde{\pi}_{\min} \in (0, 1)$.

Lemma 13 *Suppose that $\{\mathbb{G}_t\}$ is uniformly strongly connected. Then, $\Pi_{s=0}^t \hat{W}_s$ will converge to the set $\{v \mathbf{1}_N^\top \ : \ v \in \mathbb{R}^N\}$ exponentially fast as $t \to \infty$.*

Proof of Lemma 13: The lemma is a direct consequence of Theorem 2 in [6]. ∎

Lemma 14 *Suppose that $\{\mathbb{G}_t\}$ is uniformly strongly connected. Then, $(\Pi_{l=s}^t \tilde{W}_l)^{ij} = \frac{y_s^j}{y_{t+1}^i} (\Pi_{l=s}^t \hat{W}_l)^{ij}$ and $\frac{\tilde{\pi}_s^i}{y_s^i} = \frac{1}{y_s^i} \lim_{t\to\infty} (\Pi_{l=s}^t \tilde{W}_l)^{ji} = \frac{1}{N}$ for all $i, j \in \mathcal{V}$ and $s \geq 0$.*

Proof of Lemma 14: Note that for all $l \geq 0$, we have $\tilde{w}_l^{ij} = \frac{\hat{w}_l^{ij} y_l^j}{y_{l+1}^i}$. Let $\hat{W}_{s:t} = \Pi_{l=s}^t \hat{W}_l$ for all $t \geq s \geq 0$. We claim that

$$(\Pi_{l=s}^t \tilde{W}_l)^{ij} = \frac{y_s^j \hat{w}_{s:t}^{ij}}{y_{t+1}^i},$$

where $\hat{w}_{s:t}^{ij}$ is the i, j-th entry of the matrix $\hat{W}_{s:t}^{ij}$. The claim will be proved by induction on t. When $t = s + 1$,

$$\begin{aligned}
(\tilde{W}_{s+1} \tilde{W}_s)^{ij} &= \sum_{k=1}^N \tilde{w}_{s+1}^{ik} \cdot \tilde{w}_s^{kj} \\
&= \sum_{k=1}^N \frac{y_{s+1}^k \hat{w}_{s+1}^{ik}}{y_{s+2}^i} \frac{y_s^j \hat{w}_s^{kj}}{y_{s+1}^k} \\
&= \frac{y_s^j}{y_{s+2}^i} \sum_{k=1}^N \hat{w}_{s+1}^{ik} \hat{w}_s^{kj} = \frac{y_s^j}{y_{s+2}^i} \hat{w}_{s:s+1}^{ij}.
\end{aligned}$$

Thus, in this case the claim is true. Now suppose that the claim holds for all $t = \tau \geq s$, where τ is a positive integer. For $t = \tau + 1$, we have

$$\begin{aligned}
(\Pi_{s=1}^{\tau+1} \tilde{W}_s)^{ij} &= \sum_{k=1}^N \tilde{w}_{\tau+1}^{ik} \cdot \frac{y_s^j \hat{w}_{s:\tau}^{kj}}{y_{\tau+1}^k} \\
&= \sum_{k=1}^N \frac{\hat{w}_{\tau+1}^{ik} y_{\tau+1}^k}{y_{\tau+2}^i} \cdot \frac{y_s^j \hat{w}_{s:\tau}^{kj}}{y_{\tau+1}^k} \\
&= \frac{y_s^j}{y_{\tau+2}^i} \sum_{k=1}^N \hat{w}_{\tau+1}^{ik} \cdot \hat{w}_{s:\tau}^{kj} = \frac{y_s^j}{y_{\tau+2}^i} \hat{w}_{s:\tau+1}^{ij},
\end{aligned}$$

which establishes the claim by induction.

From Lemma 13, for given $s \geq 0$, we have $\lim_{t\to\infty} \hat{W}_{s:t} = v_{s,\infty}\mathbf{1}_N^\top$, with the understanding here that $v_{s,\infty}$ is not a constant vector. Then, since $y_{t+1} = \hat{W}_t y_t = \Pi_{l=s}^t \hat{W}_l y_s$ for all $t \geq s$, we have

$$\lim_{t\to\infty} (\Pi_{l=s}^t \tilde{W}_l)^{ij} = \lim_{t\to\infty} \frac{y_s^j \hat{w}_{s:t}^{ij}}{y_{t+1}^i} = \lim_{t\to\infty} \frac{y_s^j \hat{w}_{s:t}^{ij}}{\sum_{k=1}^N \hat{W}_{s:t}^{ik} y_s^k} = \frac{y_s^j \lim_{t\to\infty} \hat{w}_{s:t}^{ij}}{\lim_{t\to\infty} \sum_{k=1}^N \hat{W}_{s:t}^{ik} y_s^k}$$
$$= \frac{y_s^j v_{s,\infty}^i}{\sum_{k=1}^N v_{s,\infty}^i y_s^k} = \frac{y_s^j}{N},$$

where we use the fact that $\mathbf{1}_N^\top y_s = N$ for all $s \geq 0$ in the last equality. This completes the proof. ■

To proceed, let

$$h^j(\Theta_n, y_n) = A\frac{\theta_n^i}{y_n^i} + \frac{b^i}{y_n^i}$$
$$M_n^j = \left(A(X_n) - \mathbb{E}[A(X_n)|\mathcal{F}_{n-\tau(\alpha_n)}]\right) \frac{\theta_n^j}{y_n^j} + \frac{1}{y_n^j}\left(b^j(X_n) - \mathbb{E}[b^j(X_n)|\mathcal{F}_{n-\tau(\alpha_n)}]\right)$$
$$G_n^j = \left(\mathbb{E}[A(X_n)|\mathcal{F}_{n-\tau(\alpha_n)}] - A\right) \frac{\theta_n^j}{y_n^j} + \frac{1}{y_n^j}\left(\mathbb{E}[b^j(X_n)|\mathcal{F}_{n-\tau(\alpha_n)}] - b^j\right).$$

From (77)

$$\theta_{n+1}^i = \sum_{j=1}^N \tilde{w}_n^{ij}\left[\theta_n^j + \alpha_n h^j(\theta_n, y_n) + \alpha_n M_n^j + \alpha_n G_n^j\right].$$

Let $h = [h^1, \cdots, h^N]^\top$, $M = [M^1, \cdots, M^N]^\top$ and $G = [G^1, \cdots, G^N]^\top$. Since

$$\mathbb{E}[M_n^j|\mathcal{F}_n] = \left(\mathbb{E}[A(X_t)|\mathcal{F}_n] - \mathbb{E}[\mathbb{E}[A(X_t)|\mathcal{F}_{n-\tau(\alpha_n)}]|\mathcal{F}_n]\right)\frac{\theta_t^j}{y_t^j}$$
$$+ \frac{1}{y_t^j}\left(\mathbb{E}[b^j(X_t)|\mathcal{F}_n] - \mathbb{E}[\mathbb{E}[b^j(X_t)|\mathcal{F}_{n-\tau(\alpha_n)}]|\mathcal{F}_n]\right) = 0$$

and for all $n \geq \tau(\alpha_n)$

$$\mathbb{E}[\|M_n\|_F^2|\mathcal{F}_n] = \sum_{j=1}^N \mathbb{E}[\|M_n^j\|_2^2|\mathcal{F}_n]$$
$$= \sum_{j=1}^N \mathbb{E}[\|\left(A(X_n) - \mathbb{E}[A(X_n)|\mathcal{F}_{n-\tau(\alpha_n)}]\right)\frac{\theta_t^j}{y_t^j} + \frac{1}{y_t^j}\left(b^j(X_t) - \mathbb{E}[b^j(X_t)|\mathcal{F}_{n-\tau(\alpha_n)}]\right)\|_2^2|\mathcal{F}_n]$$
$$\leq \sum_{j=1}^N \left(\frac{2A_{\max}+\alpha_0}{\beta}\|\theta_t^j\|_2 + \frac{2b_{\max}+\alpha_0}{\beta}\right)^2 \leq \frac{2(2A_{\max}+\alpha_0)^2}{\beta^2}\|\Theta_t\|_F^2 + \frac{2N}{\beta^2}(2b_{\max}+\alpha_0)^2,$$

then $\{M_n\}$ is a martingale difference sequence satisfying $\mathbb{E}[\|M_n\|_F^2|\mathcal{F}_n] \leq \hat{C}(1 + \|\Theta_t\|_F)$, where $\hat{C} = \max\{\frac{2(2A_{\max}+\alpha_0)^2}{\beta^2}, \frac{2N}{\beta^2}(2b_{\max}+\alpha_0)^2\}$.

Define $h_c : \mathbb{R}^{N\times K} \times \mathbb{R}^N \to \mathbb{R}^{N\times K}$ as $h_c(x,y) = h(cx,y)c^{-1}$ with some $c \geq 1$. In addition, by using Lemma 14, define $\tilde{h}_c(z) : \mathbb{R}^K \to \mathbb{R}^K$ as $\tilde{h}_c(z) = h_c(\mathbf{1}_N \cdot z^\top, y_n)^\top \tilde{\pi}_n$, i.e.,

$$h_c(\Theta_n, y_n) = \begin{bmatrix} (A\frac{\theta_n^1}{y_n^1} + \frac{b^1}{y_n^1 c})^\top \\ \cdots \\ (A\frac{\theta_n^N}{y_n^N} + \frac{b^N}{y_n^N c})^\top \end{bmatrix}, \qquad \tilde{h}_c(z) = Az + \sum_{i=1}^N \frac{b^i}{Nc}.$$

Then $\tilde{h}_c(z) \to \tilde{h}_\infty(z) = Az$ as $c \to \infty$ uniformly on compact sets. Let $\phi_c(z,t)$ and $\phi_\infty(z,t)$ denote the solutions of the ODE:

$$\dot{z}(t) = \tilde{h}_c(z(t)), \quad z(0) = z \tag{80}$$
$$\dot{z}(t) = \tilde{h}_\infty(z(t)) = Az(t), \quad z(0) = z$$

respectively. Furthermore, since the origin is the unique globally asymptotically stable equilibrium of the ODE, then we have the following lemma.

Lemma 15 *There exist constant $c_0 > 0$ and $T > 0$ such that for all initial conditions z with the sphere $\{z | \|z\|_2 \le \frac{1}{N^{1/2}}\}$ and all $c \ge c_0$, we have $\|\phi_c(z,t)\|_2 < \frac{1-\kappa}{N^{1/2}}$ for $t \in [T, T+1]$ for some $0 < \kappa < 1$.*

Proof of Lemma 15: Similar to the proof of Lemma 5 in [7]. ■

Define $t_0 = 0$, $t_n = \sum_{i=0}^{n} \alpha_n$, $n \ge 0$. Define $\bar{\Theta}(t), t \ge 0$ as $\bar{\Theta}(t_n) = \Theta_n$ with linear interpolation on each interval $[t_n, t_{n+1}]$. In addition, let $T_0 = 0$ and $T_{n+1} = \min\{t_m : t_m \ge T_n + T\}$ for all $n \ge 0$. Then, $T_{n+1} \in [T_n + T, T_n + T + \sup_n \alpha_n]$. Let $m(n)$ be the value such that $T_n = t_{m(n)}$ for any $n \ge 0$. Define the piecewise continuous trajectory $\hat{\Theta}(t) = \bar{\Theta}(t) \cdot r_n^{-1}$ for $t \in [T_n, T_{n+1})$, where $r_n = \max\{\|\bar{\Theta}(T_n)\|_F, 1\}$.

Lemma 16 *There exists a positive constant $C_{\hat{\theta}} < \infty$ such that $\sup_t \|\hat{\Theta}(t)\|_F < C_{\hat{\theta}}$.*

Proof of Lemma 16: First, we write the update of $\hat{\Theta}(t_k)$ for $k \in [m(n), m(n+1))$

$$\hat{\Theta}(t_{k+1}) = \tilde{W}_{t_k} \left[\hat{\Theta}(t_k) + \alpha_{t_k} \begin{bmatrix} (\hat{\theta}^1(t_k))^\top / y^1_{t_k} \\ \cdots \\ (\hat{\theta}^N(t_k))^\top / y^N_{t_k} \end{bmatrix} A(X_{t_k})^\top + \alpha_{t_k} \begin{bmatrix} (b^1(X_{t_k}))^\top / (y^1_{t_k} r_n) \\ \cdots \\ (b^N(X_{t_k}))^\top / (y^N_{t_k} r_n) \end{bmatrix} \right]. \tag{81}$$

Since W_{t_k} is a column matrix, thus we have

$$\begin{aligned}
&\|\hat{\Theta}(t_{k+1})\|_\infty \\
&\le \|\tilde{W}_{t_k}\|_\infty \left(\|\hat{\Theta}(t_k)\|_\infty + \alpha_{t_k} \left\| \begin{bmatrix} A(X_{t_k})\hat{\theta}^1(t_k)/y^1_{t_k} \\ \cdots \\ A(X_{t_k})\hat{\theta}^N(t_k)/y^N_{t_k} \end{bmatrix} \right\|_\infty + \alpha_{t_k} \left\| \begin{bmatrix} \frac{b^1(X_{t_k})}{y^1_{t_k} r_n} \\ \cdots \\ \frac{b^N(X_{t_k})}{y^N_{t_k} r_n} \end{bmatrix} \right\|_\infty \right) \\
&\le \|\hat{\Theta}(t_k)\|_\infty + \frac{\alpha_{t_k}\sqrt{K} A_{\max}}{\beta} \|\hat{\Theta}(t_k)\|_\infty + \frac{\alpha_{t_k}\sqrt{K} b_{\max}}{\beta r_n} \\
&\le \|\hat{\Theta}(t_{m(n)})\|_\infty + \sqrt{K} \sum_{l=0}^{k-m(n)} \frac{\alpha_{t_{k+l}} A_{\max}}{\beta} \|\hat{\Theta}(t_{k+l})\|_\infty + \frac{\alpha_{t_{k+l}} b_{\max}}{\beta r_n} \\
&\le \sqrt{K} + \frac{(T + \sup_l \alpha_l)\sqrt{K} b_{\max}}{\beta} + \sum_{l=0}^{k-m(n)} \frac{\alpha_{t_{k+l}}\sqrt{K} A_{\max}}{\beta} \|\hat{\Theta}(t_{k+l})\|_\infty,
\end{aligned}$$

where we use the fact that $\|\hat{\Theta}(t_{m(n)})\|_F = 1$ and $r_n \ge 1$ in the last inequality. Therefore, by using discrete-time Grönwall inequality, we have

$$\sup_{m(n) \le k < m(n+1)} \|\hat{\Theta}(t_{k+1})\|_\infty \le \sqrt{K}(1 + (T + \sup_l \alpha_l) b_{\max}) \exp\left\{ \frac{A_{\max}\sqrt{K}}{\beta} (T + \sup_l \alpha_l) \right\}.$$

Since $T + \sup_l \alpha_l < \infty$, we have $\sup_{m(n) \le k < m(n+1)} \|\hat{\Theta}(t_{k+1})\|_\infty < \infty$ for all n. By equivalence of vector norms, we further obtain that $\sup_t \|\hat{\Theta}(t)\|_F < \infty$. ■

For $n \ge 0$, let $z^n(t)$ denote the trajectory of $\dot{z} = \tilde{h}_c(z)$ with $c = r_n$ and $z^n(T_n) = \sum_{i=1}^{N} \tilde{\pi}^i_{T_n} \hat{\theta}_{T_n}$, for $[T_n, T_{n+1})$.

882 **Lemma 17** $\lim_n \sup_{t\in[T_n,T_{n+1})} \|\hat{\Theta}_t - \mathbf{1}\otimes z^n(t)\| = 0.$

883 **Proof of Lemma 17:** From (77) and (81), for any $k \in [m(n), m(n+1))$, by Lemma 14, we have

$$\begin{aligned}
&\sum_{i=1}^{N} \tilde{\pi}_{n+1}^i \theta_{n+1}^i = \Theta_{n+1}^\top \tilde{\pi}_{n+1} \\
&= \left(\Theta_n + \alpha_n \begin{bmatrix} (A(X_n)\theta_n^1)^\top / y_n^1 \\ \cdots \\ (A(X_n)\theta_n^N)^\top / y_n^N \end{bmatrix} + \alpha_n \begin{bmatrix} (b^1(X_n))^\top / y_n^1 \\ \cdots \\ (b^N(X_n))^\top / y_n^N \end{bmatrix}\right)^\top \tilde{\pi}_n \\
&= \sum_{i=1}^{N} \tilde{\pi}_n^i \theta_n^i + \alpha_n \sum_{i=1}^{N} \tilde{\pi}_n^i (A(X_n)\theta_n^i / y_n^i + b^i(X_n)/y_n^i) \\
&= \sum_{i=1}^{N} \tilde{\pi}_n^i \theta_n^i + \frac{\alpha_n}{N} A(X_n) \sum_{i=1}^{N} \theta_n^i + \frac{\alpha_n}{N} \sum_{i=1}^{N} b^i(X_n).
\end{aligned}$$

884 Similarly, we have

$$\begin{aligned}
&\sum_{i=1}^{N} \tilde{\pi}_{t_{k+1}}^i \hat{\theta}_{t_{k+1}}^i \\
&= \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i + \alpha_t \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i (A(X_{t_k})\hat{\theta}_{t_k}^i / y_{t_k}^i + b^i(X_{t_k})/(y_{t_k}^i r_n)) \\
&= \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i + \alpha_{t_k} \left(A(X_{t_k}) \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i + \frac{1}{N r_n} \sum_{i=1}^{N} b^i(X_{t_k}) \right) \\
&\quad + \alpha_{t_k} \frac{A(X_{t_k})}{N} \sum_{i=1}^{N} \left(\hat{\theta}_{t_k}^i - \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i \right) \\
&= \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i + \alpha_{t_k} \left(A \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i + \frac{1}{N r_n} \sum_{i=1}^{N} b^i \right) + \alpha_{t_k} \frac{A(X_{t_k})}{N} \sum_{i=1}^{N} \left(\hat{\theta}_{t_k}^i - \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i \right) \\
&\quad + \alpha_{t_k} \left(A(X_{t_k}) - \mathbb{E}[A(X_{t_k}) | \mathcal{F}_{t_k - \tau(\alpha_{t_k})}] \right) \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i \\
&\quad + \frac{\alpha_{t_k}}{N r_n} \sum_{i=1}^{N} \left(b^i(X_{t_k}) - \mathbb{E}[b^i(X_{t_k}) | \mathcal{F}_{t_k - \tau(\alpha_{t_k})}] \right) \\
&\quad + \alpha_{t_k} \left(\left(\mathbb{E}[A(X_{t_k}) | \mathcal{F}_{t_k - \tau(\alpha_{t_k})}] - A \right) \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i + \frac{1}{N r_n} \sum_{i=1}^{N} \left(\mathbb{E}[b^i(X_{t_k}) | \mathcal{F}_{t_k - \tau(\alpha_{t_k})}] - b^i \right) \right).
\end{aligned}$$

885 Let

$$\begin{aligned}
\hat{M}_{t_k} &= \left(A(X_t) - \mathbb{E}[A(X_{t_k}) | \mathcal{F}_{t_k - \tau(\alpha_{t_k})}] \right) \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i \\
&\quad + \frac{1}{N r_n} \sum_{i=1}^{N} \left(b^i(X_{t_k}) - \mathbb{E}[b^i(X_{t_k}) | \mathcal{F}_{t_k - \tau(\alpha_{t_k})}] \right) \\
\hat{G}_{t_k} &= \left(\mathbb{E}[A(X_{t_k}) | \mathcal{F}_{t_k - \tau(\alpha_{t_k})}] - A \right) \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i + \frac{1}{N r_n} \sum_{i=1}^{N} \left(\mathbb{E}[b^i(X_{t_k}) | \mathcal{F}_{t_k - \tau(\alpha_{t_k})}] - b^i \right) \\
&\quad + \frac{A(X_{t_k})}{N} \sum_{i=1}^{N} \left(\hat{\theta}_{t_k}^i - \sum_{i=1}^{N} \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i \right).
\end{aligned}$$

It is easy to verify that $\{\hat{M}_{t_k}\}$ is a martingale difference sequence satisfying $\mathbb{E}[\|\hat{M}_{t_k}\|_2^2|\mathcal{F}_{t_k}] \leq \bar{C}(1+\|\sum_{i=1}^N \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i\|_2^2)$ for some $\bar{C} \leq \infty$. In addition, we have

$$\begin{aligned}
&\hat{\theta}_{t_k}^i - \sum_{j=1}^N \tilde{\pi}_{t_k}^j \hat{\theta}_{t_k}^j \\
&= \sum_{j=1}^N (\tilde{w}_{t_s:t_k}^{ij} - \tilde{\pi}_{t_s}^j)\hat{\theta}_{t_s}^j + \sum_{r=s+1}^k \alpha_{t_r} \sum_{i=1}^N (\tilde{w}_{t_r:t_k}^{ij} - \tilde{\pi}_{t_r}^j)(A(X_{t_r})\hat{\theta}_{t_r}^j / y_{t_r}^j + b^j(X_{t_r}/y_{t_r}^j).
\end{aligned}$$

Since $\{\mathbb{G}_t\}$ is uniformly strongly connected, then for any $s \geq 0$, $W_{s:t}$ converges to $\mathbf{1}\pi_s^\top$ exponentially fast as $t \to \infty$ and there exist a finite positive constant C and a constant $0 \leq \lambda < 1$ such that

$$|\tilde{w}_{s:t}^{ij} - \tilde{\pi}_s^j| \leq C\lambda^{t-s}$$

for all $i, j \in \mathcal{V}$ and $s \geq 0$. Then,for any $k \in [m(n), m(n+1))$, we have

$$\begin{aligned}
&\|\hat{\theta}_{t_k}^i - \sum_{j=1}^N \tilde{\pi}_{t_k}^j \hat{\theta}_{t_k}^j\|_2 \\
&\leq \sum_{j=1}^N \|\tilde{w}_{t_{m(n)}:t_k}^{ij} - \tilde{\pi}_{t_{m(n)}}^j\|_2 \|\hat{\theta}_{t_{m(n)}}^j\|_2 + \sum_{r=m(n)+1}^k \alpha_{t_r} \sum_{i=1}^N \|\tilde{w}_{t_r:t_k}^{ij} - \tilde{\pi}_{t_r}^j\|_2 \frac{A_{\max}\|\hat{\theta}_{t_r}^j\|_2 + b_{\max}}{\beta} \\
&\leq \sum_{j=1}^N C\lambda^{t_k - t_{m(n)}} \|\hat{\theta}_{t_{m(n)}}^j\|_2 + \sum_{r=m(n)+1}^k \alpha_{t_r} \sum_{i=1}^N C\lambda^{t_k - t_r} (\frac{A_{\max}\|\hat{\theta}_{t_r}^j\|_2 + b_{\max}}{\beta}) \\
&\leq NC\lambda^{t_k - t_{m(n)}} + \frac{\alpha_{t_{m(n)}} NC}{1-\lambda} \frac{A_{\max} C_{\hat{\theta}} + b_{\max}}{\beta},
\end{aligned}$$

where in the last inequality, we use the fact that for all $n \geq 0$, we have $\|\hat{\Theta}(t_{m(n)})\|_F = 1$, $\alpha_{n+1} \leq \alpha_n$ and the boundedness of $\|\hat{\Theta}_n\|_F$ from Lemma 16. Since $\alpha_{t_k} \to 0$ as $k \to \infty$, then

$$\lim_{k\to\infty} \|\hat{\theta}_{t_k}^i - \sum_{j=1}^N \tilde{\pi}_{t_k}^j \hat{\theta}_{t_k}^j\|_2 = 0,$$

which implies that

$$\lim_{k\to\infty} \left\| \frac{A(X_{t_k})}{N} \sum_{i=1}^N (\hat{\theta}_{t_k}^i - \sum_{j=1}^N \tilde{\pi}_{t_k}^j \hat{\theta}_{t_k}^j) \right\|_2 = 0.$$

Then,

$$\lim_{k\to\infty} \|\hat{G}_{t_k}\|_2 \leq \lim_{k\to\infty} \alpha_{t_k}(\|\sum_{j=1}^N \tilde{\pi}_{t_k}^j \hat{\theta}_{t_k}^j\|_2 + 1) + \lim_{k\to\infty} \left\| \frac{A(X_{t_k})}{N} \sum_{i=1}^N (\hat{\theta}_{t_k}^i - \sum_{j=1}^N \tilde{\pi}_{t_k}^j \hat{\theta}_{t_k}^j) \right\|_2 = 0.$$

Therefore, by Corollary 8 and Theorem 9 in Chapter 6 of [8], we obtain that $\sum_{i=1}^N \tilde{\pi}_{t_k}^i \hat{\theta}_{t_k}^i \to z^n(t)$ as $n \to \infty$, namely $k \to \infty$. Furthermore, we obtain that $\hat{\theta}_{t_{k+1}}^i \to z^n(t)$ as $n \to \infty$ for all $i \in \mathcal{V}$, which concludes the proof following Theorem 2 in Chapter 2 of [8]. ■

Lemma 18 *The sequence $\{\Theta_n\}$ generated from* (79) *is bounded almost surely, i.e., $C_\theta = \sup_n \|\Theta_n\|_F < \infty$ almost surely.*

Proof of Lemma 18: In order to prove this lemma, we need to show that $\sup_n \|\bar{\Theta}(T_n)\|_F < \infty$ first. If this does not hold, there will exist a sequence $T_{n_1}, T_{n_2}, \cdots$ such that $\|\hat{\Theta}(T_{n_k})\|_F \to \infty$, i.e., $r_{n_k} \to \infty$. If $r_n > c_0$ and $\|\hat{\Theta}(T_n)\|_F = 1$, then $\|z^n(T_n)\|_2 = \|\sum_{i=1}^N \tilde{\pi}_{T_n} \hat{\theta}_{T_n}^i\|_2 \leq N^{-1/2}$. Using Lemma 15, we have $\|\mathbf{1}_N \cdot (z^n(T_{n+1}^-))^\top\|_F = N^{1/2}\|z^n(T_{n+1}^-)\|_2 \leq 1-\kappa$. In addition, using

Lemma 17, there exists a constant $0 < \kappa' < \kappa$ such that $\|\hat{\Theta}(T_{n+1}^-)\|_F < 1 - \kappa'$. Hence for $r_n > c_0$ and n sufficiently large,

$$\frac{\|\bar{\Theta}(T_{n+1})\|_F}{\|\bar{\Theta}(T_n)\|_F} = \frac{\|\hat{\Theta}(T_{n+1}^-)\|_F}{\|\hat{\Theta}(T_n)\|_F} \leq 1 - \kappa'.$$

It shows that if $\|\bar{\Theta}(T_n)\|_F > c_0$, $\|\bar{\Theta}(T_k)\|_F$ for all $k \geq n$ falls back to the ball of radius c_0 at an exponential rate.

Thus, if $\|\bar{\Theta}(T_n)\|_F > c_0$, then $\|\bar{\Theta}(T_{n-1})\|_F$ is either greater than $\|\bar{\Theta}(T_n)\|_F$ or is inside the ball of radius c_0. Since we assume $r_{n_k} \to \infty$, then we can find a time T_n such that $\|\bar{\Theta}(T_n)\|_F < c_0$ and $\|\bar{\Theta}(T_{n+1})\|_F = \infty$. However, by using discrete-time Grönwall inequality, we have

$$\begin{aligned}
\|\bar{\Theta}(T_{n+1})\|_\infty &\leq \|\bar{\Theta}(T_{n+1}-1)\|_\infty + \alpha_{T_{n+1}-1}\frac{\sqrt{K}A_{\max}}{\beta}\|\bar{\Theta}(T_{n+1}-1)\|_\infty + \alpha_{T_{n+1}-1}\sqrt{K}\frac{b_{\max}}{\beta} \\
&\leq \|\bar{\Theta}(T_n)\|_\infty + \sqrt{K}\sum_{s=0}^{T_{n+1}-T_n} \alpha_{T_n+s}\frac{A_{\max}}{\beta}\|\bar{\Theta}(T_n+s)\|_\infty + \alpha_{T_n+s}\frac{b_{\max}}{\beta} \\
&\leq \sqrt{K}c_0 + \sqrt{K}(T + \sup_n \alpha_n)\frac{b_{\max}}{\beta} + \frac{\sqrt{K}A_{\max}}{\beta}\sum_{s=0}^{T_{n+1}-T_n} \alpha_{T_n+s}\|\bar{\Theta}(T_n+s)\|_\infty \\
&\leq \sqrt{K}(c_0 + (T + \sup_n \alpha_n)\frac{b_{\max}}{\beta}) \exp\left\{(T + \sup_n \alpha_n)\frac{\sqrt{K}A_{\max}}{\beta}\right\},
\end{aligned}$$

which implies that $\|\bar{\Theta}(T_{n+1})\|_F$ can be bounded if $\|\bar{\Theta}(T_n)\|_F < c_0$. This leads to a contradiction.

Moreover, let $C_{\bar{\theta}} = \sup_n \|\bar{\Theta}(T_n)\|_F < \infty$, then $C_\theta = \sup_n \|\Theta_n\|_F \leq C_{\bar{\theta}} C_{\hat{\theta}} < \infty$. ■

Recall the update of $\tilde{\theta}_t^i$ in (9):

$$\tilde{\theta}_{t+1}^i = \sum_{j=1}^N \hat{w}_t^{ij}\left[\tilde{\theta}_t^j + \alpha_t\left(A(X_t)\theta_t^j + b^j(X_t)\right)\right].$$

From the definition that $\langle\tilde{\theta}\rangle_t = \frac{1}{N}\sum_{i=1}^N \tilde{\theta}_t^i$ and $\langle\theta\rangle_t = \frac{1}{N}\sum_{i=1}^N \theta_t^i$, we have

$$\begin{aligned}
\langle\tilde{\theta}\rangle_{t+1} &= \langle\tilde{\theta}\rangle_t + \alpha_t A(X_t)\langle\theta\rangle_t + \frac{\alpha_t}{N}\sum_{i=1}^N b^i(X_t) \\
&= \langle\tilde{\theta}\rangle_t + \alpha_t A(X_t)\langle\tilde{\theta}\rangle_t + \frac{\alpha_t}{N}\sum_{i=1}^N b^i(X_t) + \alpha_t\rho_t, \qquad (82)
\end{aligned}$$

where $\rho_t = A(X_t)\langle\theta\rangle_t - A(X_t)\langle\tilde{\theta}\rangle_t$. From Lemma 18, we have $\|\langle\theta\rangle_t\|_2 \leq \max_{i\in\mathcal{V}}\|\theta_t^i\|_2 \leq C_\theta$ for all $t \geq 0$, which implies that $\|\langle\tilde{\theta}\rangle_t\|_2 \leq NC_\theta$ and

$$\mu_t = \|\rho_t\|_2 = \left\|A(X_t)\langle\theta\rangle_t - A(X_t)\langle\tilde{\theta}\rangle_t\right\|_2 \leq \mu_{\max},$$

where $\mu_{\max} = (N+1)A_{\max}C_\theta$.

Lemma 19 *Suppose that Assumptions 2 and 5 hold and $\{\mathbb{G}_t\}$ is uniformly strongly connected by sub-sequences of length L. Let $\epsilon_1 = \inf_{t\geq 0}\min_{i\in\mathcal{V}}(\hat{W}_t\cdots\hat{W}_0\mathbf{1}_N)^i$. For all $t \geq 0$ and $i \in \mathcal{V}$,*

$$\begin{aligned}
\|\theta_{t+1}^i - \langle\tilde{\theta}\rangle_t\|_2 \leq{}& \frac{8}{\epsilon_1}\bar{\epsilon}^t\|\sum_{i=1}^N \tilde{\theta}_0^i + \alpha_0 A(X_0)\theta_0^i + \alpha_0 b^i(X_0)\|_2 \\
&+ \frac{8}{\epsilon_1}\frac{A_{\max}C_\theta + b_{\max}}{1-\bar{\epsilon}}\left(\alpha_0\bar{\epsilon}^{t/2} + \alpha_{\lceil\frac{t}{2}\rceil}\right) + \alpha_t A_{\max}C_\theta + \alpha_t b_{\max},
\end{aligned}$$

where $\epsilon_1 > 0$ and $\bar{\epsilon} \in (0,1)$ satisfy $\epsilon_1 \geq \frac{1}{N^{NL}}$ and $\bar{\epsilon} \leq (1 - \frac{1}{N^{NL}})^{1/L}$.

Proof of Lemma 19: Since $\epsilon_1 = \inf_{t\geq 0}\min_{i\in\mathcal{V}}(\hat{W}_t\cdots\hat{W}_0\mathbf{1}_N)^i$ and all weight matrices $\hat{W}_s$ are column stochastic matrices for all $s\geq 0$, from Corollary 2 (b) in [1], we know that $\epsilon_1 \leq \frac{1}{N^{NL}}$. If the weight matrices are doubly stochastic matrices, then $\epsilon_1 = 1$.

From Assumption 2 and Lemma 18, we know that $\|A(X_t)\theta_t^i + b^i(X_t)\|_2 \leq A_{\max}C_\theta + b_{\max}$. Then, by using Lemma 1 in [1], for all $t\geq 0$ and $i\in\mathcal{V}$ we have

$$
\begin{aligned}
&\|\theta_{t+1}^i - \langle\tilde{\theta}\rangle_t - \alpha_t A(X_t)\langle\theta\rangle_t - \frac{\alpha_t}{N}\sum_{i=1}^N b^i(X_t)\|_2 \\
&\leq \frac{8}{\epsilon_1}(\bar{\epsilon}^t\|\sum_{i=1}^N\tilde{\theta}_0^i + \alpha_0 A(X_0)\theta_0^i + \alpha_0 b^i(X_0)\|_2 + \sum_{s=0}^t \bar{\epsilon}^{t-s}\alpha_s(A_{\max}C_\theta + b_{\max})) \\
&\leq \frac{8}{\epsilon_1}\bar{\epsilon}^t\|\sum_{i=1}^N\tilde{\theta}_0^i + \alpha_0 A(X_0)\theta_0^i + \alpha_0 b^i(X_0)\|_2 \\
&\quad + \frac{8}{\epsilon_1}(A_{\max}C_\theta + b_{\max})\left(\sum_{s=0}^{\lfloor\frac{t}{2}\rfloor}\bar{\epsilon}^{t-s}\alpha_s + \sum_{s=\lceil\frac{t}{2}\rceil}^t \bar{\epsilon}^{t-s}\alpha_s\right) \\
&\leq \frac{8}{\epsilon_1}\bar{\epsilon}^t\|\sum_{i=1}^N\tilde{\theta}_0^i + \alpha_0 A(X_0)\theta_0^i + \alpha_0 b^i(X_0)\|_2 + \frac{8}{\epsilon_1}\frac{A_{\max}C_\theta + b_{\max}}{1-\bar{\epsilon}}\left(\alpha_0\bar{\epsilon}^{t/2} + \alpha_{\lceil\frac{t}{2}\rceil}\right),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
&\|\theta_{t+1}^i - \langle\tilde{\theta}\rangle_t\|_2 \\
&\leq \|\theta_{t+1}^i - \langle\tilde{\theta}\rangle_t - \alpha_t A(X_t)\langle\theta\rangle_t - \frac{\alpha_t}{N}\sum_{i=1}^N b^i(X_t)\|_2 + \alpha_t\|A(X_t)\langle\theta\rangle_t + \frac{1}{N}\sum_{i=1}^N b^i(X_t)\|_2 \\
&\leq \frac{8}{\epsilon_1}\bar{\epsilon}^t\|\sum_{i=1}^N\tilde{\theta}_0^i + \alpha_0 A(X_0)\theta_0^i + \alpha_0 b^i(X_0)\|_2 + \frac{8}{\epsilon_1}\frac{A_{\max}C_\theta + b_{\max}}{1-\bar{\epsilon}}\left(\alpha_0\bar{\epsilon}^{t/2} + \alpha_{\lceil\frac{t}{2}\rceil}\right) \\
&\quad + \alpha_t A_{\max}C_\theta + \alpha_t b_{\max}.
\end{aligned}
$$

This completes the proof. ■

Lemma 20 $\lim_{t\to\infty}\mu_t = \lim_{t\to\infty}\|\rho_t\|_2 = 0$ *and* $\lim_{t\to\infty}\frac{\sum_{k=0}^t\mu_k}{t+1} = \lim_{t\to\infty}\frac{\sum_{k=0}^t\|\rho_k\|_2}{t+1} = 0$.

Proof of Lemma 20: From Lemma 19, we have

$$
\begin{aligned}
\mu_t = \|\rho_t\|_2 &= \left\|A(X_t)\langle\theta\rangle_t - A(X_t)\langle\tilde{\theta}\rangle_t\right\|_2 \\
&\leq \frac{8A_{\max}}{\epsilon_1}\bar{\epsilon}^t\|\tilde{\Theta}_0\|_1 + \frac{8A_{\max}}{\epsilon_1}\frac{N\sqrt{K}(A_{\max}C_\theta + b_{\max})}{1-\bar{\epsilon}}\left(\alpha_0\bar{\epsilon}^{t/2} + \alpha_{\lceil\frac{t}{2}\rceil}\right).
\end{aligned}
$$

Since $\bar{\epsilon}\in(0,1)$, then $\lim_{t\to\infty}\|\rho_t\|_2 = 0$. Next, we will prove that $\lim_{t\to\infty}\frac{1}{t+1}\sum_{k=0}^t\|\rho_k\|_2 = 0$. For any positive constant $c>0$, there exists a positive integer $T(c)$, depending on c, such that $\forall t\geq T(c)$, we have $\|\rho_t\|_2 < c$. Thus,

$$
\frac{1}{t}\sum_{k=0}^{t-1}\|\rho_k\|_2 = \frac{1}{t}\sum_{k=0}^{T(c)}\|\rho_k\|_2 + \frac{1}{t}\sum_{k=T(c)+1}^{t-1}\|\rho_k\|_2 \leq \frac{1}{t}\sum_{k=0}^{T(c)}\|\rho_k\|_2 + \frac{t-1-T(c)}{t}c.
$$

Let $t\to\infty$ on both sides of the above inequality. Then, we have

$$
\lim_{t\to\infty}\frac{1}{t}\sum_{k=0}^{t-1}\|\rho_k\|_2 \leq \lim_{t\to\infty}\frac{1}{t}\sum_{k=0}^{T(c)}\|\rho_k\|_2 + \lim_{t\to\infty}\frac{t-1-T(c)}{t}c = c.
$$

Since the above argument holds for arbitrary positive c, then $\lim_{t\to\infty}\frac{1}{t+1}\sum_{k=0}^t\|\rho_k\|_2 = 0$. ■

Lemma 21 *Suppose that Assumptions 2 and 3 hold. When the step-size α_t and corresponding mixing time $\tau(\alpha_t)$ satisfy $0 < \alpha_t \tau(\alpha_t) < \frac{\log 2}{A_{\max}}$, we have for any $t \geq \bar{T}$,*

$$\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \leq 2A_{\max}\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k + 2(b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k, \tag{83}$$

$$\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \leq 6A_{\max}\|\langle\tilde{\theta}\rangle_t\|_2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k + 5(b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k, \tag{84}$$

$$\begin{aligned}\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 &\leq 72\alpha_{t-\tau(\alpha_t)}^2 \tau^2(\alpha_t) A_{\max}^2 \|\langle\tilde{\theta}\rangle_t\|_2^2 + 50\alpha_{t-\tau(\alpha_t)}^2 \tau^2(\alpha_t)(b_{\max} + \mu_{\max})^2 \\ &\leq 8\|\langle\tilde{\theta}\rangle_t\|_2^2 + \frac{6(b_{\max} + \mu_{\max})^2}{A_{\max}^2}. \end{aligned} \tag{85}$$

Proof of Lemma 21: From the update of $\langle\tilde{\theta}\rangle_t$ in (82):

$$\langle\tilde{\theta}\rangle_{t+1} = \langle\tilde{\theta}\rangle_t + \alpha_t A(X_t)\langle\tilde{\theta}\rangle_t + \frac{\alpha_t}{N}\sum_{i=1}^{N} b^i(X_t) + \alpha_t \rho_t.$$

Then, we have

$$\|\langle\tilde{\theta}\rangle_{t+1}\|_2 \leq (1 + \alpha_t A_{\max})\|\langle\tilde{\theta}\rangle_t\|_2 + \alpha_t b_{\max} + \alpha_t \mu_{\max}.$$

For all $u \in [t - \tau(\alpha_t), t]$, we have

$$\begin{aligned}\|\langle\tilde{\theta}\rangle_u\|_2 &\leq \Pi_{k=t-\tau(\alpha_t)}^{u-1}(1 + \alpha_k A_{\max})\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \\ &\quad + (b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{u-1} \alpha_k \Pi_{l=k+1}^{u-1}(1 + \alpha_l A_{\max}) \\ &\leq \exp\{\sum_{k=t-\tau(\alpha_t)}^{u-1} \alpha_k A_{\max}\}\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \\ &\quad + (b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{u-1} \alpha_k \exp\{\sum_{l=k+1}^{u-1} \alpha_l A_{\max}\} \\ &\leq \exp\{\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)A_{\max}\}\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \\ &\quad + (b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{u-1} \alpha_k \exp\{\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)A_{\max}\} \\ &\leq 2\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 + 2(b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{u-1} \alpha_k, \end{aligned}$$

where we use $\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)A_{\max} \le \log 2 < \frac{1}{3}$ in the last inequality. Thus, for all $t \ge \bar{T}$, we can get (83) as follows:

$$
\begin{aligned}
&\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \\
&\le \sum_{k=t-\tau(\alpha_t)}^{t-1} \|\langle\tilde{\theta}\rangle_{k+1} - \langle\tilde{\theta}\rangle_k\|_2 \\
&\le A_{\max} \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \|\langle\tilde{\theta}\rangle_k\|_2 + (b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le A_{\max} \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \left(2\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 + 2(b_{\max} + \mu_{\max}) \sum_{l=t-\tau(\alpha_t)}^{k-1} \alpha_l \right) \\
&\quad + (b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 2A_{\max}\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k + \left(2A_{\max}\tau(\alpha_t)\alpha_{t-\tau(\alpha_t)} + 1\right)(b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 2A_{\max}\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k + \frac{5}{3}(b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 2A_{\max}\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k + 2(b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k .
\end{aligned}
$$

Moreover, by using the above inequality, we can get (84) for all $t \ge \bar{T}$ as follows:

$$
\begin{aligned}
&\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \\
&\le 2A_{\max}\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k + \frac{5}{3}(b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 2A_{\max}\tau(\alpha_t)\alpha_{t-\tau(\alpha_t)}\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 + 2A_{\max}\|\langle\tilde{\theta}\rangle_t\|_2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + \frac{5}{3}(b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 6A_{\max}\|\langle\tilde{\theta}\rangle_t\|_2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k + 5(b_{\max} + \mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k .
\end{aligned}
$$

Next, by using (84) and the inequality $(x+y)^2 \le 2x^2 + y^2$ for all x, y, we can get (85) as follows:

$$
\begin{aligned}
\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 &\le 72A_{\max}^2\|\langle\tilde{\theta}\rangle_t\|_2^2 \Big(\sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k\Big)^2 + 50(b_{\max} + \mu_{\max})^2 \Big(\sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k\Big)^2 \\
&\le 72\alpha_{t-\tau(\alpha_t)}^2\tau^2(\alpha_t)A_{\max}^2\|\langle\tilde{\theta}\rangle_t\|_2^2 + 50\alpha_{t-\tau(\alpha_t)}^2\tau^2(\alpha_t)(b_{\max} + \mu_{\max})^2 \\
&\le 8\|\langle\tilde{\theta}\rangle_t\|_2^2 + \frac{6(b_{\max} + \mu_{\max})^2}{A_{\max}^2},
\end{aligned}
$$

where we use $\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)A_{\max} < \frac{1}{3}$ in the last inequality. ∎

Lemma 22 *Suppose that Assumptions 2–5 hold and $\{\mathbb{G}_t\}$ is uniformly strongly connected by sub-sequences of length L. When $0 < \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t) < \frac{\log 2}{A_{\max}}$, we have for any $t \geq \bar{T}$,*

$$
\begin{aligned}
&|\mathbf{E}[(\langle\tilde{\theta}\rangle_t - \theta^*)^\top (P+P^\top)(A(X_t)\langle\tilde{\theta}\rangle_t + B(X_t)^\top \pi_{t+1} - A\langle\tilde{\theta}\rangle_t - b) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\leq \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\left(72 + 456A_{\max}^2 + 84A_{\max}b_{\max} + 72A_{\max}\mu_{\max}\right)\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\
&\quad + \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\Bigg[2 + 4\|\theta^*\|_2^2 + 48\frac{(b_{\max}+\mu_{\max})^2}{A_{\max}^2} + 152\left(b_{\max}+\mu_{\max}+A_{\max}\|\theta^*\|_2\right)^2 \\
&\quad + 12A_{\max}b_{\max} + 48A_{\max}(b_{\max}+\mu_{\max})(\frac{b_{\max}+\mu_{\max}}{A_{\max}}+1)^2 + 87(b_{\max}+\mu_{\max})^2\Bigg].
\end{aligned}
$$

Proof of Lemma 22: Note that for all $t \geq \bar{T}$, we have

$$
\begin{aligned}
&\mathbf{E}[(\langle\tilde{\theta}\rangle_t - \theta^*)^\top (P+P^\top)(A(X_t)\langle\tilde{\theta}\rangle_t + \frac{1}{N}B(X_t)^\top \mathbf{1}_N - A\langle\tilde{\theta}\rangle_t - b) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\leq |\mathbf{E}[(\langle\tilde{\theta}\rangle_t - \theta^*)^\top (P+P^\top)(A(X_t)-A)\langle\tilde{\theta}\rangle_t \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\quad + |\mathbf{E}[(\langle\tilde{\theta}\rangle_t - \theta^*)^\top (P+P^\top)(\frac{1}{N}B(X_t)^\top \mathbf{1}_N - b) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\leq |\mathbf{E}[(\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top (P+P^\top)(A(X_t)-A)\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} \mid \mathcal{F}_{t-\tau(\alpha_t)}]| && (86)\\
&\quad + |\mathbf{E}[(\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top (P+P^\top)(A(X_t)-A)(\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| && (87)\\
&\quad + |\mathbf{E}[(\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)})^\top (P+P^\top)(A(X_t)-A)\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} \mid \mathcal{F}_{t-\tau(\alpha_t)}]| && (88)\\
&\quad + |\mathbf{E}[(\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)})^\top (P+P^\top)(A(X_t)-A)(\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| && (89)\\
&\quad + |\mathbf{E}[(\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)})^\top (P+P^\top)(\frac{1}{N}B(X_t)^\top \mathbf{1}_N - b) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| && (90)\\
&\quad + |\mathbf{E}[(\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top (P+P^\top)(\frac{1}{N}B(X_t)^\top \mathbf{1}_N - b) \mid \mathcal{F}_{t-\tau(\alpha_t)}]|. && (91)
\end{aligned}
$$

By using the mixing time in Assumption 3, we can get the bound for (86) and (91) for all $t \geq \bar{T}$:

$$
\begin{aligned}
&|\mathbf{E}[(\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top (P+P^\top)(A(X_t)-A)\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\leq |(\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top (P+P^\top)\mathbf{E}[A(X_t)-A \mid \mathcal{F}_{t-\tau(\alpha_t)}]\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}| \\
&\leq 2\alpha_t\gamma_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \theta^*\|_2\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\
&\leq \alpha_t\gamma_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \theta^*\|_2^2 + \|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\
&\leq \alpha_t\gamma_{\max}\mathbf{E}[2\|\theta^*\|_2^2 + 3\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\
&\leq 6\alpha_t\gamma_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] + 6\alpha_t\gamma_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] + 2\alpha_t\gamma_{\max}\|\theta^*\|_2^2 \\
&\leq 54\alpha_t\gamma_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] + 36\alpha_t\gamma_{\max}\frac{(b_{\max}+\mu_{\max})^2}{A_{\max}^2} + 2\alpha_t\gamma_{\max}\|\theta^*\|_2^2, && (92)
\end{aligned}
$$

where in the last inequality, we use (83) from Lemma 21.

$$
\begin{aligned}
&|\mathbf{E}[(\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top (P+P^\top)(\frac{1}{N}B(X_t)^\top \mathbf{1}_N - b) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\leq |(\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top (P+P^\top)\frac{1}{N}\sum_{i=1}^N \mathbf{E}[b^i(X_t) - b^i \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\leq 2\gamma_{\max}\alpha_t\mathbf{E}[\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \theta^*\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\
&\leq 2\gamma_{\max}\alpha_t\left(1 + \frac{1}{2}\mathbf{E}[\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] + \frac{1}{2}\|\theta^*\|_2^2\right) \\
&\leq 2\gamma_{\max}\alpha_t\left(1 + \mathbf{E}[\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \langle\tilde{\theta}\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] + \mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] + \|\theta^*\|_2^2\right) \\
&\leq 2\gamma_{\max}\alpha_t\left(1 + 9\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] + 6\frac{(b_{\max}+\mu_{\max})^2}{A_{\max}^2} + \|\theta^*\|_2^2\right), && (93)
\end{aligned}
$$

950 where in the last inequality we use (83).

951 Next, by using Assumption 2, (83) and (85), we have

$$
\begin{aligned}
&|\mathbf{E}[(\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \theta^*)^\top (P + P^\top)(A(X_t) - A)(\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le 4\gamma_{\max} A_{\max} \mathbf{E}[\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} - \theta^*\|_2 \|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\
&\le 4\gamma_{\max} A_{\max} \mathbf{E}[\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\
&\quad + 4\gamma_{\max} A_{\max} \|\theta^*\|_2 \mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\
&\le 8\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 8\gamma_{\max} A_{\max} (b_{\max} + \mu_{\max}) \|\theta^*\|_2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 8\gamma_{\max} A_{\max}^2 \left(\frac{b_{\max} + \mu_{\max}}{A_{\max}} + \|\theta^*\|_2 \right) \mathbf{E}[\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 12\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 8\gamma_{\max} A_{\max}^2 \left(\frac{b_{\max} + \mu_{\max}}{A_{\max}} + \|\theta^*\|_2 \right)^2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 24\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 24\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 8\gamma_{\max} A_{\max}^2 \left(\frac{b_{\max} + \mu_{\max}}{A_{\max}} + \|\theta^*\|_2 \right)^2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 216\gamma_{\max} A_{\max}^2 \mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 152\gamma_{\max} \left(b_{\max} + \mu_{\max} + A_{\max} \|\theta^*\|_2 \right)^2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k.
\end{aligned} \tag{94}
$$

952 In additional, as for the bound of (88), by using (83) and (85), we have

$$
\begin{aligned}
&|\mathbf{E}[(\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)})^\top (P+P^\top)(A(X_t)-A)\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)} \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le 4\gamma_{\max}A_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\
&\le 8\gamma_{\max}A_{\max}\mathbf{E}[A_{\max}\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 + (b_{\max}+\mu_{\max})\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 4\gamma_{\max}A_{\max}(2A_{\max}+b_{\max}+\mu_{\max})\mathbf{E}[\|\langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 4\gamma_{\max}A_{\max}(b_{\max}+\mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 8\gamma_{\max}A_{\max}(2A_{\max}+b_{\max}+\mu_{\max})\mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 8\gamma_{\max}A_{\max}(2A_{\max}+b_{\max}+\mu_{\max})\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 4\gamma_{\max}A_{\max}(b_{\max}+\mu_{\max}) \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 72\,\gamma_{\max}A_{\max}(2A_{\max}+b_{\max}+\mu_{\max})\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + 48\gamma_{\max}A_{\max}(b_{\max}+\mu_{\max})(\frac{b_{\max}+\mu_{\max}}{A_{\max}}+1)^2 \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k.
\end{aligned} \tag{95}
$$

953 Moreover, by using (85), we can get the bound for (89) as follows:

$$
\begin{aligned}
&|\mathbf{E}[(\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)})^\top (P+P^\top)(A(X_t)-A)(\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le 4\gamma_{\max}A_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le 4\gamma_{\max}A_{\max}\mathbf{E}[72A_{\max}^2\|\langle\tilde{\theta}\rangle_t\|_2^2 + 50(b_{\max}+\mu_{\max})^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \left(\sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k\right)^2 \\
&\le 96A_{\max}^2\gamma_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k + 67(b_{\max}+\mu_{\max})^2\gamma_{\max} \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k.
\end{aligned} \tag{96}
$$

954 Finally, we can get the bound of (90) by using (84):

$$
\begin{aligned}
&|\mathbf{E}[(\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)})(P+P^\top)(\frac{1}{N}B(X_t)^\top \mathbf{1}_N - b) \mid \mathcal{F}_{t-\tau(\alpha_t)}]| \\
&\le 4\gamma_{\max}b_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \langle\tilde{\theta}\rangle_{t-\tau(\alpha_t)}\|_2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \\
&\le 4\gamma_{\max}b_{\max}\mathbf{E}[6A_{\max}\|\langle\tilde{\theta}\rangle_t\|_2 + 5(b_{\max}+\mu_{\max}) \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\le 12\gamma_{\max}A_{\max}b_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2 \mid \mathcal{F}_{t-\tau(\alpha_t)}] \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k \\
&\quad + (12A_{\max}+20b_{\max}+20\mu_{\max})\gamma_{\max}b_{\max} \sum_{k=t-\tau(\alpha_t)}^{t-1} \alpha_k.
\end{aligned} \tag{97}
$$

Then, by using (92)–(97), we have

$$
\begin{aligned}
&|\mathbf{E}[(\langle\tilde{\theta}\rangle_t-\theta^*)^\top(P+P^\top)(A(X_t)\langle\tilde{\theta}\rangle_t+B(X_t)^\top\pi_{t+1}-A\langle\tilde{\theta}\rangle_t-b)\mid\mathcal{F}_{t-\tau(\alpha_t)}]|\\
&\le 54\alpha_t\gamma_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]+36\alpha_t\gamma_{\max}\frac{(b_{\max}+\mu_{\max})^2}{A_{\max}^2}+2\alpha_t\gamma_{\max}\|\theta^*\|_2^2\\
&\quad+2\gamma_{\max}\alpha_t\left(1+9\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]+6\frac{(b_{\max}+\mu_{\max})^2}{A_{\max}^2}+\|\theta^*\|_2^2\right)\\
&\quad+216\gamma_{\max}A_{\max}^2\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k\\
&\quad+152\gamma_{\max}\left(b_{\max}+\mu_{\max}+A_{\max}\|\theta^*\|_2\right)^2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k\\
&\quad+72\,\gamma_{\max}A_{\max}(2A_{\max}+b_{\max}+\mu_{\max})\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k\\
&\quad+48\gamma_{\max}A_{\max}(b_{\max}+\mu_{\max})(\frac{b_{\max}+\mu_{\max}}{A_{\max}}+1)^2\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k\\
&\quad+96A_{\max}^2\gamma_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k+67(b_{\max}+\mu_{\max})^2\gamma_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k\\
&\quad+12\gamma_{\max}A_{\max}b_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k\\
&\quad+(12A_{\max}+20b_{\max}+20\mu_{\max})\gamma_{\max}b_{\max}\sum_{k=t-\tau(\alpha_t)}^{t-1}\alpha_k\\
&\le\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\left(72+456A_{\max}^2+84A_{\max}b_{\max}+72A_{\max}\mu_{\max}\right)\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2\mid\mathcal{F}_{t-\tau(\alpha_t)}]\\
&\quad+\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\Bigg[2+4\|\theta^*\|_2^2+48\frac{(b_{\max}+\mu_{\max})^2}{A_{\max}^2}+152\left(b_{\max}+\mu_{\max}+A_{\max}\|\theta^*\|_2\right)^2\\
&\quad+12A_{\max}b_{\max}+48A_{\max}(b_{\max}+\mu_{\max})(\frac{b_{\max}+\mu_{\max}}{A_{\max}}+1)^2+87(b_{\max}+\mu_{\max})^2\Bigg],
\end{aligned}
$$

where we use $\alpha_t\le\alpha_{t-\tau\alpha_t}$ from Assumption 5 and $\tau(\alpha_t)\ge 1$ in the last inequality. This completes the proof. ■

Lemma 23 *Suppose that Assumptions 2–4 hold and* $\alpha_t=\frac{\alpha_0}{t+1}$. *When* $\mu_t+\tau(\alpha_t)\alpha_{t-\tau(\alpha_t)}\zeta_8\le\frac{0.1}{\gamma_{\max}}$ *and* $\tau(\alpha_t)\alpha_{t-\tau(\alpha_t)}\le\min\{\frac{\log 2}{A_{\max}},\ \frac{0.1}{\zeta_8\gamma_{\max}}\}$, *we have for* $t\ge\bar{T}$,

$$
\begin{aligned}
\mathbf{E}[\|\langle\tilde{\theta}\rangle_{t+1}-\theta^*\|_2^2]\le\frac{\bar{T}}{t+1}\frac{\gamma_{\max}}{\gamma_{\min}}\mathbf{E}[\|\langle\tilde{\theta}\rangle_{\bar{T}}-\theta^*\|_2^2]+\frac{\zeta_9\alpha_0C\log^2(\frac{t+1}{\alpha_0})}{t+1}\frac{\gamma_{\max}}{\gamma_{\min}}\\
+\alpha_0\frac{\gamma_{\max}}{\gamma_{\min}}\frac{\sum_{l=\bar{T}}^{t+1}\mu_l}{t+1},
\end{aligned}
$$

where $\bar{T}$ *is defined in Appendix A.2,* ζ_8 *and* ζ_9 *are defined in* (18) *and* (19)*, respectively.*

Proof of Lemma 23: Let $H(\langle\tilde{\theta}\rangle_t)=(\langle\tilde{\theta}\rangle_t-\theta^*)^\top P(\langle\tilde{\theta}\rangle_t-\theta^*)$. From Assumption 4, we know that

$$\gamma_{\min}\|\langle\tilde{\theta}\rangle_t-\theta^*\|_2^2\le H(\langle\tilde{\theta}\rangle_t)\le\gamma_{\max}\|\langle\tilde{\theta}\rangle_t-\theta^*\|_2^2.$$

Recall the update of $\langle\tilde{\theta}\rangle_t$ in (82):

$$\langle\tilde{\theta}\rangle_{t+1}=\langle\tilde{\theta}\rangle_t+\alpha_tA(X_t)\langle\tilde{\theta}\rangle_t+\frac{\alpha_t}{N}\sum_{i=1}^{N}b^i(X_t)+\alpha_t\rho_t.$$

From Assumption 2, for $t \geq \bar{T}$ we have

$$
\begin{aligned}
&H(\langle\tilde{\theta}\rangle_{t+1}) \\
&= (\langle\tilde{\theta}\rangle_{t+1} - \theta^*)^\top P(\langle\tilde{\theta}\rangle_{t+1} - \theta^*) \\
&= \left(\langle\tilde{\theta}\rangle_t + \alpha_t A(X_t)\langle\tilde{\theta}\rangle_t + \frac{\alpha_t}{N}\sum_{i=1}^N b^i(X_t) + \alpha_t\rho_t - \theta^*\right)^\top P\cdot \\
&\quad \left(\langle\tilde{\theta}\rangle_t + \alpha_t A(X_t)\langle\tilde{\theta}\rangle_t + \frac{\alpha_t}{N}\sum_{i=1}^N b^i(X_t) + \alpha_t\rho_t - \theta^*\right) \\
&= (\langle\tilde{\theta}\rangle_t - \theta^*)^\top P(\langle\tilde{\theta}\rangle_t - \theta^*) + \alpha_t^2(A(X_t)\langle\tilde{\theta}\rangle_t)^\top P(A(X_t)\langle\tilde{\theta}\rangle_t) \\
&\quad + \frac{\alpha_t^2}{N^2}(B(X_t)^\top\mathbf{1}_N)^\top P(B(X_t)^\top\mathbf{1}_N) + \frac{\alpha_t^2}{N}(A(X_t)\langle\tilde{\theta}\rangle_t)^\top(P+P^\top)(B(X_t)^\top\mathbf{1}_N) + \alpha_t^2\rho_t^\top P\rho_t \\
&\quad + \alpha_t^2(A(X_t)\langle\tilde{\theta}\rangle_t + \frac{1}{N}B(X_t)^\top\mathbf{1}_N)^\top(P+P^\top)\rho_t + \alpha_t(\langle\tilde{\theta}\rangle_t - \theta^*)^\top(P+P^\top)\rho_t \\
&\quad + \alpha_t(\langle\tilde{\theta}\rangle_t - \theta^*)^\top(P+P^\top)(A(X_t)\langle\tilde{\theta}\rangle_t + \frac{1}{N}B(X_t)^\top\mathbf{1}_N - A\langle\tilde{\theta}\rangle_t - b) \\
&\quad + \alpha_t(\langle\tilde{\theta}\rangle_t - \theta^*)^\top P(A\langle\tilde{\theta}\rangle_t + b) + \alpha_t(A\langle\tilde{\theta}\rangle_t + b)^\top P(\langle\tilde{\theta}\rangle_t - \theta^*) \\
&= H(\langle\tilde{\theta}\rangle_t) + \alpha_t^2(A(X_t)\langle\tilde{\theta}\rangle_t)^\top P(A(X_t)\langle\tilde{\theta}\rangle_t) + \frac{\alpha_t^2}{N^2}(B(X_t)^\top\mathbf{1}_N)^\top P(B(X_t)^\top\mathbf{1}_N) \\
&\quad + \frac{\alpha_t^2}{N}(A(X_t)\langle\tilde{\theta}\rangle_t)^\top(P+P^\top)(B(X_t)^\top\mathbf{1}_N) + \alpha_t^2\rho_t^\top P\rho_t \\
&\quad + \alpha_t^2(A(X_t)\langle\tilde{\theta}\rangle_t + \frac{1}{N}B(X_t)^\top\mathbf{1}_N)^\top(P+P^\top)\rho_t + \alpha_t(\langle\tilde{\theta}\rangle_t - \theta^*)^\top(P+P^\top)\rho_t \\
&\quad + \alpha_t(\langle\tilde{\theta}\rangle_t - \theta^*)^\top(P+P^\top)(A(X_t)\langle\tilde{\theta}\rangle_t + \frac{1}{N}B(X_t)^\top\mathbf{1}_N - A\langle\tilde{\theta}\rangle_t - b) \\
&\quad + \alpha_t(\langle\tilde{\theta}\rangle_t - \theta^*)^\top(PA + A^\top P)(\langle\tilde{\theta}\rangle_t - \theta^*),
\end{aligned} \tag{98}
$$

where we use the fact that $A\theta^* + b = 0$ on the last equality.

965 Next, we can take expectation on both sides of (98). From Assumption 4 and Lemma 22, for $t \geq \bar{T}$
966 we have

$$
\begin{aligned}
&\mathbf{E}[H(\langle\tilde{\theta}\rangle_{t+1})] \\
&= \mathbf{E}[H(\langle\tilde{\theta}\rangle_t)] + \alpha_t^2 \mathbf{E}[(A(X_t)\langle\tilde{\theta}\rangle_t)^\top P(A(X_t)\langle\tilde{\theta}\rangle_t)] - \alpha_t \mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \theta^*\|_2^2] + \mathbf{E}[\alpha_t^2 \rho_t^\top P \rho_t] \\
&\quad + \frac{\alpha_t^2}{N^2}\mathbf{E}[(B(X_t)^\top \mathbf{1}_N)^\top P (B(X_t)^\top \mathbf{1}_N)] + \frac{\alpha_t^2}{N}\mathbf{E}[(A(X_t)\langle\tilde{\theta}\rangle_t)^\top (P + P^\top)(B(X_t)^\top \mathbf{1}_N)] \\
&\quad + \alpha_t^2 \mathbf{E}[(A(X_t)\langle\tilde{\theta}\rangle_t + \frac{1}{N}B(X_t)^\top \mathbf{1}_N)^\top (P + P^\top)\rho_t] + \alpha_t \mathbf{E}[(\langle\tilde{\theta}\rangle_t - \theta^*)^\top (P + P^\top)\rho_t] \\
&\quad + \alpha_t \mathbf{E}[(\langle\tilde{\theta}\rangle_t - \theta^*)^\top (P + P^\top)(A(X_t)\langle\tilde{\theta}\rangle_t + \frac{1}{N}B(X_t)^\top \mathbf{1}_N - A\langle\tilde{\theta}\rangle_t - b)] \\
&\leq \mathbf{E}[H(\langle\tilde{\theta}\rangle_t)] + \alpha_t^2 A_{\max}^2 \gamma_{\max} \mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2] - \alpha_t \mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \theta^*\|_2^2] + 2\alpha_t \gamma_{\max}\|\rho_t\|_2 \mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \theta^*\|_2] \\
&\quad + \alpha_t^2 \gamma_{\max} b_{\max}^2 + 2\alpha_t^2 \gamma_{\max} A_{\max} b_{\max} \mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2] + \alpha_t^2 \gamma_{\max}\mu_{\max}^2 \\
&\quad + 2\alpha_t^2 \gamma_{\max}\mu_{\max}(A_{\max}\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2] + b_{\max}) \\
&\quad + \alpha_t \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\left(72 + 456 A_{\max}^2 + 84 A_{\max} b_{\max} + 72 A_{\max}\mu_{\max}\right)\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2] \\
&\quad + \alpha_t \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\Bigg[2 + 48\frac{(b_{\max}+\mu_{\max})^2}{A_{\max}^2} + 152\left(b_{\max} + \mu_{\max} + A_{\max}\|\theta^*\|_2\right)^2 \\
&\quad + 4\|\theta^*\|_2^2 + 12 A_{\max} b_{\max} + 48 A_{\max}(b_{\max}+\mu_{\max})(\frac{b_{\max}+\mu_{\max}}{A_{\max}} + 1)^2 + 87(b_{\max}+\mu_{\max})^2\Bigg] \\
&\leq \mathbf{E}[H(\langle\tilde{\theta}\rangle_t)] + (-\alpha_t + \alpha_t\gamma_{\max}\|\rho_t\|_2)\mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \theta^*\|_2^2] + \alpha_t \gamma_{\max}\|\rho_t\|_2 \\
&\quad + \alpha_t \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\left(72 + 458 A_{\max}^2 + 84 A_{\max} b_{\max} + 72 A_{\max}\mu_{\max}\right)\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2] \\
&\quad + \alpha_t \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\Bigg[2 + 48\frac{(b_{\max}+\mu_{\max})^2}{A_{\max}^2} + 152\left(b_{\max} + \mu_{\max} + A_{\max}\|\theta^*\|_2\right)^2 \\
&\quad + 4\|\theta^*\|_2^2 + 12 A_{\max} b_{\max} + 48 A_{\max}(b_{\max}+\mu_{\max})(\frac{b_{\max}+\mu_{\max}}{A_{\max}} + 1)^2 + 89(b_{\max}+\mu_{\max})^2\Bigg].
\end{aligned}
$$

967 Using the facts that $\mathbf{E}[\|\langle\tilde{\theta}\rangle_t\|_2^2] \leq 2\mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \theta^*\|_2^2] + 2\|\theta^*\|_2^2$ and $\gamma_{\min}\|\langle\tilde{\theta}\rangle_t - \theta^*\|_2^2 \leq H(\langle\tilde{\theta}\rangle_t) \leq$
968 $\gamma_{\max}\|\langle\tilde{\theta}\rangle_t - \theta^*\|_2^2$, then

$$
\begin{aligned}
&\mathbf{E}[H(\langle\tilde{\theta}\rangle_{t+1})] \\
&\leq \mathbf{E}[H(\langle\tilde{\theta}\rangle_t)] + (-\alpha_t + \alpha_t\gamma_{\max}\mu_t)\mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \theta^*\|_2^2] + \alpha_t\gamma_{\max}\mu_t \\
&\quad + 2\alpha_t^2\gamma_{\max}(b_{\max}+\mu_{\max})^2 \\
&\quad + 2\alpha_t \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\left(72 + 458 A_{\max}^2 + 84 A_{\max} b_{\max} + 72 A_{\max}\mu_{\max}\right)\mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \theta^*\|_2^2] \\
&\quad + 2\alpha_t \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\left(72 + 458 A_{\max}^2 + 84 A_{\max} b_{\max} + 72 A_{\max}\mu_{\max}\right)\|\theta^*\|_2^2 \\
&\quad + \alpha_t \alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\Bigg[2 + 48\frac{(b_{\max}+\mu_{\max})^2}{A_{\max}^2} + 152\left(b_{\max} + \mu_{\max} + A_{\max}\|\theta^*\|_2\right)^2 \\
&\quad + 4\|\theta^*\|_2^2 + 12 A_{\max} b_{\max} + 48 A_{\max}(b_{\max}+\mu_{\max})(\frac{b_{\max}+\mu_{\max}}{A_{\max}} + 1)^2 + 87(b_{\max}+\mu_{\max})^2\Bigg] \\
&\leq \mathbf{E}[H(\langle\tilde{\theta}\rangle_t)] + (-\alpha_t + \alpha_t\gamma_{\max}\mu_t + \alpha_t\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\zeta_8)\mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \theta^*\|_2^2] \\
&\quad + \alpha_t\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\zeta_9 + \alpha_t\gamma_{\max}\mu_t.
\end{aligned}
$$

Moreover, from $\alpha_t = \frac{\alpha_0}{t+1}$, $\alpha_0 \geq \frac{\gamma_{max}}{0.9}$ and the definition of $\bar{T}$, for all $t \geq \bar{T}$ we have

$$\begin{aligned}
\mathbf{E}[H(\langle\tilde{\theta}\rangle_{t+1})] &\leq (1 - \frac{0.9\alpha_t}{\gamma_{\max}})\mathbf{E}[H(\langle\tilde{\theta}\rangle_t)] + \alpha_t\alpha_{t-\tau(\alpha_t)}\tau(\alpha_t)\gamma_{\max}\zeta_9 + \alpha_t\gamma_{\max}\mu_t \\
&\leq \frac{t}{t+1}\mathbf{E}[H(\langle\tilde{\theta}\rangle_t)] + \alpha_0\gamma_{\max}\frac{\mu_t}{t+1} + \frac{\alpha_0^2 C\log(\frac{t+1}{\alpha_0})\gamma_{\max}\zeta_9}{(t+1)(t-\tau(\alpha_t)+1)} \\
&\leq \frac{\bar{T}}{t+1}\mathbf{E}[H(\langle\tilde{\theta}\rangle_{\bar{T}})] + \alpha_0\gamma_{\max}\sum_{l=\bar{T}}^{t}\frac{\mu_l}{l+1}\Pi_{u=l+1}^{t}\frac{u}{u+1} \\
&\quad + \alpha_0^2\gamma_{\max}\zeta_9\sum_{l=\bar{T}}^{t}\frac{C\log(\frac{l+1}{\alpha_0})}{(l+1)(l-\tau(\alpha_l)+1)}\Pi_{u=l+1}^{t}\frac{u}{u+1} \\
&= \frac{\bar{T}}{t+1}\mathbf{E}[H(\langle\tilde{\theta}\rangle_{\bar{T}})] + \alpha_0\gamma_{\max}\sum_{l=\bar{T}}^{t}\frac{\mu_l}{t+1} + \frac{\alpha_0^2\gamma_{\max}\zeta_9}{t+1}\sum_{l=\bar{T}}^{t}\frac{C\log(\frac{l+1}{\alpha_0})}{l-\tau(\alpha_l)+1} \\
&\leq \frac{\bar{T}}{t+1}\mathbf{E}[H(\langle\tilde{\theta}\rangle_{\bar{T}})] + \alpha_0\gamma_{\max}\frac{\sum_{l=\bar{T}}^{t}\mu_l}{t+1} + \frac{\alpha_0^2\gamma_{\max}\zeta_9}{t+1}\sum_{l=\bar{T}}^{t}\frac{2C\log(\frac{l+1}{\alpha_0})}{l+1} \\
&\leq \frac{\bar{T}}{t+1}\mathbf{E}[H(\langle\tilde{\theta}\rangle_{\bar{T}})] + \alpha_0\gamma_{\max}\frac{\sum_{l=\bar{T}}^{t+1}\mu_l}{t+1} + \frac{\zeta_9\alpha_0\gamma_{\max}C\log^2(\frac{t+1}{\alpha_0})}{t+1},
\end{aligned} \tag{99}$$

where we use

$$\sum_{l=\bar{T}}^{t}\frac{2\alpha_0\log(\frac{l+1}{\alpha_0})}{l+1} \leq \log^2(\frac{t+1}{\alpha_0})$$

to get the last inequality. Then, we can get the bound of $\mathbf{E}[\|\langle\tilde{\theta}\rangle_{t+1} - \theta^*\|_2^2]$ from (99) as follows:

$$\begin{aligned}
&\mathbf{E}[\|\langle\tilde{\theta}\rangle_{t+1} - \theta^*\|_2^2] \\
&\leq \frac{1}{\gamma_{\min}}\mathbf{E}[H(\langle\tilde{\theta}\rangle_{t+1})] \\
&\leq \frac{\bar{T}}{t+1}\frac{\gamma_{\max}}{\gamma_{\min}}\mathbf{E}[\|\langle\tilde{\theta}\rangle_{\bar{T}} - \theta^*\|_2^2] + \frac{\zeta_9\alpha_0 C\log^2(\frac{t+1}{\alpha_0})}{t+1}\frac{\gamma_{\max}}{\gamma_{\min}} + \alpha_0\frac{\gamma_{\max}}{\gamma_{\min}}\frac{\sum_{l=\bar{T}}^{t+1}\mu_l}{t+1}.
\end{aligned}$$

This completes the proof. ∎

We are now in a position to prove Theorem 5.

Proof of Theorem 5: Note that

$$\sum_{i=1}^{N}\mathbf{E}[\|\theta_{t+1}^i - \theta^*\|_2^2] \leq 2\sum_{i=1}^{N}\mathbf{E}[\|\theta_{t+1}^i - \langle\tilde{\theta}\rangle_t\|_2^2] + 2N\mathbf{E}[\|\langle\tilde{\theta}\rangle_t - \theta^*\|_2^2].$$

By using Lemmas 19 and 23, for any $t \geq \bar{T}$, we have

$$\begin{aligned}
&\sum_{i=1}^{N}\mathbf{E}\left[\left\|\theta_{t+1}^i - \theta^*\right\|_2^2\right] \\
&\leq \frac{16}{\epsilon_1}\bar{\epsilon}^t\mathbf{E}[\|\sum_{i=1}^{N}\tilde{\theta}_0^i + \alpha_0 A(X_0)\theta_0^i + \alpha_0 b^i(X_0)\|_2] + \frac{16}{\epsilon_1}\frac{A_{\max}C_\theta + b_{\max}}{1-\bar{\epsilon}}\left(\alpha_0\bar{\epsilon}^{t/2} + \alpha_{\lceil\frac{t}{2}\rceil}\right) \\
&\quad + 2\alpha_t A_{\max}C_\theta + 2\alpha_t b_{\max} + \frac{2\bar{T}N}{t}\frac{\gamma_{\max}}{\gamma_{\min}}\mathbf{E}[\|\langle\tilde{\theta}\rangle_{\bar{T}} - \theta^*\|_2^2] \\
&\quad + \frac{2N\zeta_9\alpha_0 C\log^2(\frac{t}{\alpha_0})}{t}\frac{\gamma_{\max}}{\gamma_{\min}} + 2\alpha_0 N\frac{\gamma_{\max}}{\gamma_{\min}}\frac{\sum_{l=\bar{T}}^{t}\mu_l}{t} \\
&\leq C_7\bar{\epsilon}^t + C_8\left(\alpha_0\bar{\epsilon}^{\frac{t}{2}} + \alpha_{\lceil\frac{t}{2}\rceil}\right) + C_9\alpha_t + \frac{1}{t}\left(C_{10}\log^2\left(\frac{t}{\alpha_0}\right) + C_{11}\sum_{l=\bar{T}}^{t}\mu_l + C_{12}\right).
\end{aligned}$$

This completes the proof. ∎

References

[1] A. Nedić and A. Olshevsky. Distributed optimization over time-varying directed graphs. *IEEE Transactions on Automatic Control*, 60(3):601–615, 2015.

[2] B. Touri. *Product of Random Stochastic Matrices and Distributed Averaging*. Springer Science & Business Media, 2012.

[3] A. Nedić and J. Liu. On convergence rate of weighted-averaging dynamics for consensus problems. *IEEE Transactions on Automatic Control*, 62(2):766–781, 2017.

[4] H.J. Kushner and G. Yin. Asymptotic properties of distributed and communicating stochastic approximation algorithms. *SIAM Journal on Control and Optimization*, 25(5):1266–1290, 1987.

[5] H. J. Kushner. An averaging method for stochastic approximations with discontinuous dynamics, constraints, and state dependent noise. In *Recent advances in statistics*, pages 211–235. Elsevier, 1983.

[6] J. Hajnal and M. S. Bartlett. Weak ergodicity in non-homogeneous Markov chains. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54:233–246, 1958.

[7] A. S. Mathkar and V. S. Borkar. Nonlinear gossip. *SIAM Journal on Control and Optimization*, 54(3):1535–1557, 2016.

[8] V. S. Borkar. *Stochastic approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.